CALCULATION

of

CONFORMATIONAL

ENERGIES

for

CARBOHYDRATES

PETER J C SMITH BSc

DOCTOR of PHILOSOPHY

UNIVERSITY of EDINBURGH

1972

PREFACE.

This thesis describes work carried out by me under the supervision of Professor D.A. Rees in the Department of Chemistry of the University of Edinburgh between October 1969 and September 1971, and on leave of absence at the Unilever Research Laboratory, Colworth House, in Bedfordshire, from October 1971 until September 1972.

I would like to express my most sincere thanks to Prof. Rees for introducing me to the field of biophysical science and especially for his constant guidance and encouragement over the last three years.

I am also very grateful to Drs. C.T. Greenwood and M.A.D. Fluendy, who acted as internal supervisors, to the Professors and the Department of Chemistry at Edinburgh and to Unilever Research at Colworth for providing research facilities, to the ever-helpful staff of the Edinburgh Regional Computing Centre, to Marine Colloids Inc. and Unilever Research for maintenance grants, and particularly to many friends and colleagues in Edinburgh and at Colworth for help, advice, support and encouragement.

September 1972.

NOTE

Almost all work published in this field is in c.g.s. units, a convention which has been continued in this thesis. The relevant conversion factors are: 1 kcal = 4.2 kJ, 1 Å = 0.1 nm.

ABSTRACT

Physical and calculation methods of determining conformation in carbohydrates are reviewed.

An intramolecular potential function is formulated, incorporating terms for non-bonded van der Waals interactions, polar interactions, the bond torsion effect and hydrogen bonding. Methods of calculating thermodynamic average energies are discussed. The effects of molecular environment and of entropy contributions to conformational free energy are considered.

The conformational enthalpies of pentoses are calculated and compared with experiment. Moderate correlation is demonstrated.

The effect of the conformation of disaccharide linkages is considered, and a method is described and applied for the calculation thence of helix parameters. The conformational enthalpy surface in terms of linkage rotation angles is calculated for maltose and cellobiose. The results are correlated with known crystal solution and polysaccharide structures, and show better agreement than previously published work, although fewer assumptions regarding specific hydrogen bonds are made.

Some polysaccharide applications are described. The geometry of cation coordination to alginates and pectin is investigated and a model proposed. The conformational enthalpy difference between coil and double helix forms of carrageenan is calculated, and other terms discussed in relation to the observed cooperativity of the transition. It is concluded that desolvation of the coil is an important effect, and that the observed optical rotation change is due to the displacement of the linkage conformation of the helix from the position of minimum linkage conformational energy. The relative stabilities of the right- and left-handed structures for triple helical β-1,3 xylan are calculated and found to be almost equal. The single and double helical proposed structures of B-amylose are briefly discussed.

CONTENTS

Appendices.

CHAPTER 1

INTRODUCTION -

THE DETERMINATION OF CARBOHYDRATE CONFORMATION.

1.1 PREAMBLE

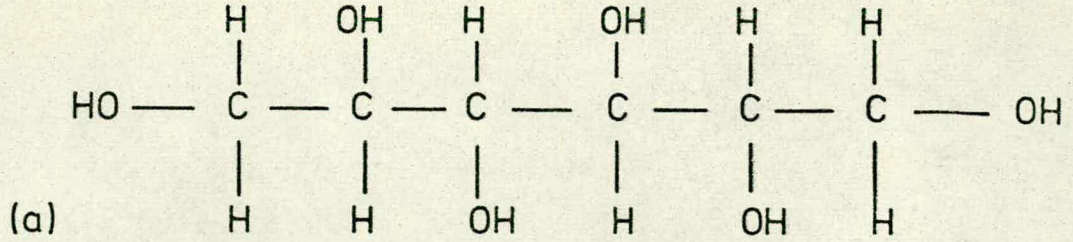
The principal aim of physical science is to describe and thereby predict the behaviour of matter: that of biology to understand the observed characteristics of living things. The advent of computers has enabled scientists in many fields to formulate such descriptions in formal mathematical terms: to describe physical, chemical and biological properties on an atomic or molecular basis.

This work is an attempt to further the study of the shapes adopted by carbohydrate molecules. Carbohydrates as sugars and polymers are vital constituents of all living systems, and form a considerable and essential part of the diet of all animals, including humans. A full knowledge of their structure and an understanding of the factors determining it, is thus of both immediate and long term importance.

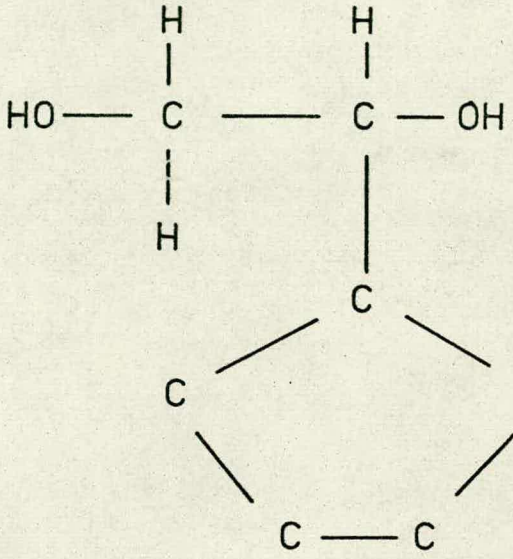
1.2 THE CONFORMATIONAL PROBLEMS IN CARBOHYDRATES

Once the chemical constitution of a monosaccharide has been determined, there remain three areas of structural uncertainty.

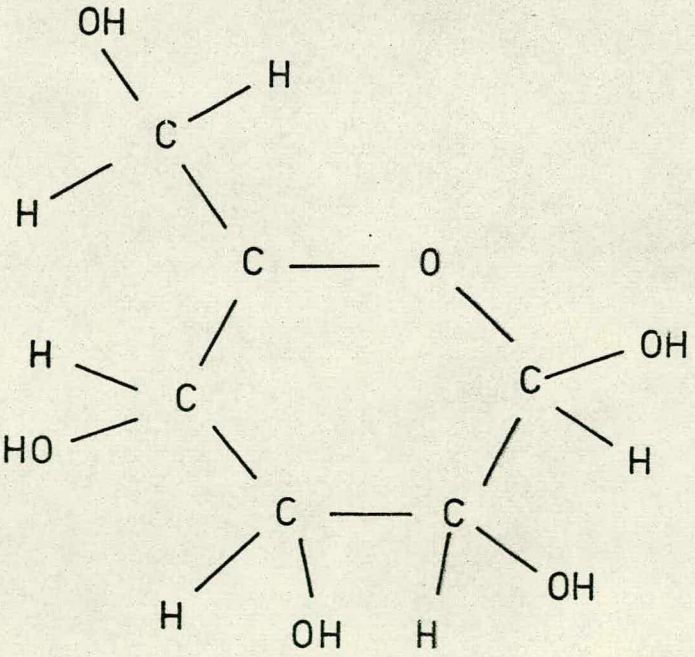
Firstly, there is the question of the arrangement of the carbon backbone of the molecule. In the molecules to be considered, pentoses and hexoses, is the predominant form (fig.1a) the 'straight' chain, or (b) the five-



(a)



(b)



(c)

Fig. 1. Possible structure for a hexose: (a) straight chain, (b) furanose ring, (c) pyranose ring.

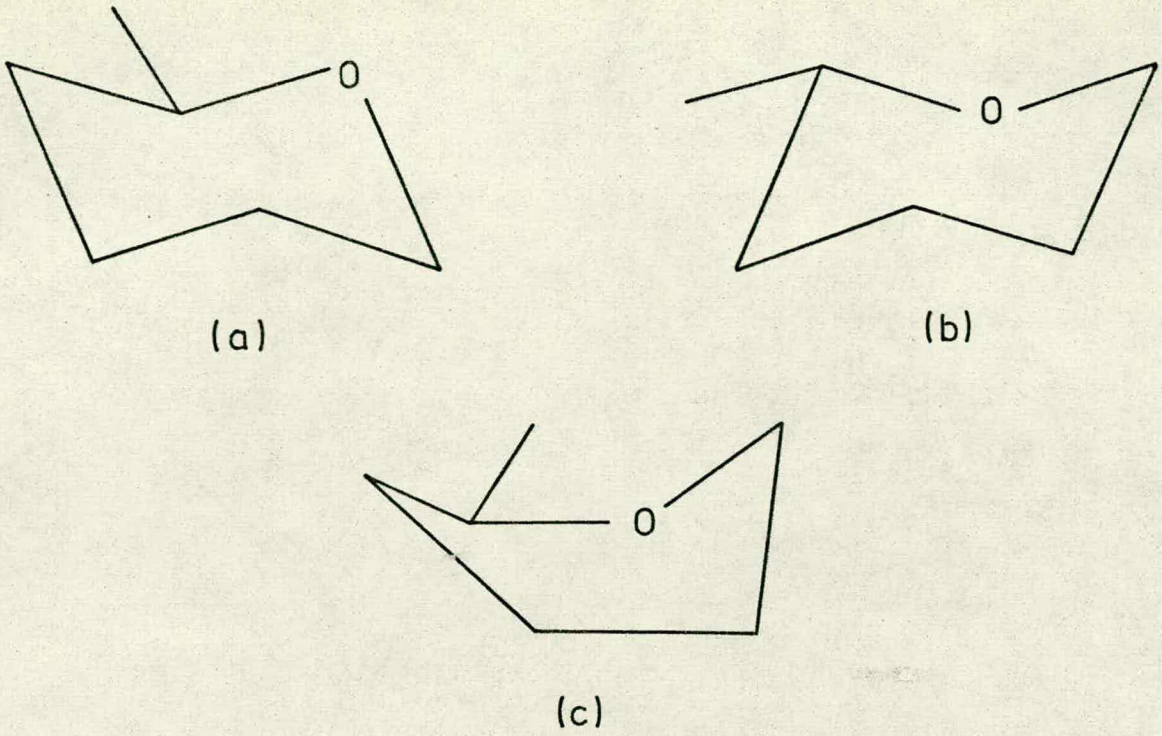


Fig. 2. Possible conformations of the pyranose ring:
(a) Reeves C1 chair, (b) Reeves 1C chair,
(c) Reeves B1 boat (there exist several other boats).

membered (furanose)ring, or (c) the six-membered (pyranose) ring, or some other form, or even a mixture? It is, in fact well established that most such sugars exist mainly in the pyranose form and this will be the only one to be considered in this thesis. Within the class of pyranose rings, there can exist without undue straining of tetrahedral bond angles several shapes of the ring (fig. 2) of which the most important and prevalent [1] are

(a) The two 'chair' conformations, and

(b) The several 'boat' forms.

Fortunately the potential minima in which these exist have been shown to be sharp and steep-sided, so that to a good approximation any given sugar exists in one or other

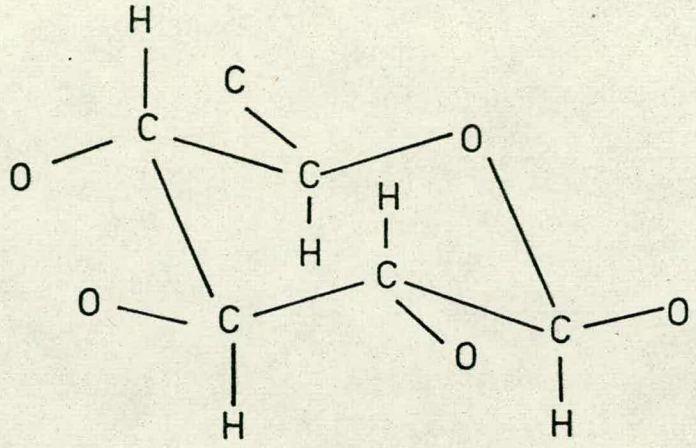


Fig. 3. For a specified ring conformation, the positions of these atoms are all fixed by the assumption of rigid bonds and bond angles.

(or possibly a mixture) of these conformations, and not in any intermediate form. Partly because of this, the ring conformation of most sugars has been established for some time.

If the positions of the ring atoms are thus fixed, certain other atoms (fig. 3) are also fixed in position merely by assuming fixed bond lengths and angles. There remain, however, other 'pendant' atoms which can 'float' by rotation about single bonds. A second problem, therefore, is the disposition of these atoms, and this is discussed in later sections of this thesis.

Thirdly, as a special case of the pendant group problem, consideration must be given to the situation in

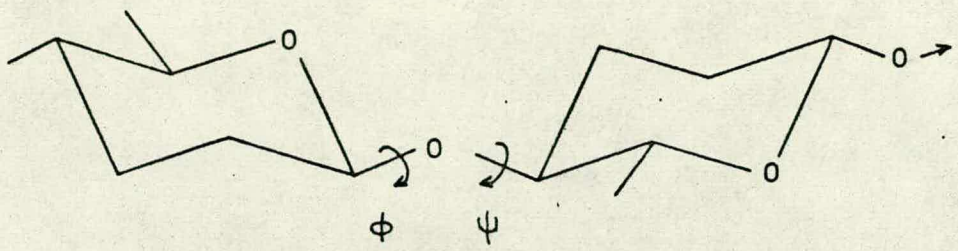


Fig. 4. Angles ϕ and ψ determining linkage or chain conformation.

higher sugars where the pendant group is another sugar ring. This situation may be viewed more symmetrically (fig. 4) as involving two degrees of freedom, namely rotation about the two angles designated ϕ and ψ. These constitute the 'linkage conformation'. In a given polysaccharide the linkage conformation may be the same (within narrow limits) between each pair of residues, or it may vary more or less randomly. This behaviour will determine the shape of the polymer, its 'chain conformation'. Regular linkage conformation gives a helical chain, randomly distributed linkages give a flexible molecule, the 'random coil'.

This last property is one which plays a major role in the determination of the physical properties of the polysaccharides, and conversely, in order to understand the structural properties of these important molecules, a knowledge of the contributory molecular properties is essential.

1.3. METHODS OF CONFORMATIONAL ANALYSIS

The determination of conformation may be achieved by a variety of experimental methods, or alternatively, predictions can be made from several calculation methods based on more or less empirical assumptions.

Experimental methods, described in sections 1.4 to 1.8 include equilibrium constant determination and their thermodynamic analysis and systematisation, nuclear magnetic resonance (NMR), x-ray diffraction, chiroptical properties and infra-red spectroscopy. With the exception of x-ray diffraction, these generally provide information on the position of equilibrium between different conformational or configurational states. X-ray work can, in principle, yield the complete structure of any crystallisable substance, and in the carbohydrate field has been thus successfully applied to many monosaccharides and smaller oligosaccharides. It is also of great use in the examination of the one-dimensional 'crystallinity' in helical polysaccharides.

From a theoretical standpoint, the methods of

quantum mechanics may be used completely to describe the properties of a molecule of known structure, in particular its relative energy and hence its preferred conformation. In practice, however, for molecules as large as carbohydrates, this method is prohibitively costly in computing time. Nevertheless, simplified treatments have had some success. The others of the calculation methods discussed in sections 1.9 to 1.12 are more empirically based: they assume more or less classical behaviour of nonbonded atoms. Although inherently less exact, these methods have had notable triumphs.

1.4 EXPERIMENT-BASED WORK - EARLY METHODS.

Although it was Haworth [2] writing in 1929 who is generally credited with introducing the term 'conformation', it was not until about 1950 that conformational analysis in the sense now accepted began to be developed. In that year Barton [3] published a paper on the conformation of the steroid nucleus (fig. 5), which is often cited [4] as the forerunner of modern work. In this and subsequent work for which he was awarded the Nobel Prize in 1969, Barton is concerned with the determination of the relative stabilities of equatorial and axial (then called polar) configurations of the pendant groups (fig. 6) on single and fused cyclohexane rings. Reviewing the literature in 1956, Barton and Cookson [5] described how thermodynamic concepts had been successfully applied to a wide range of single and fused ring compounds.

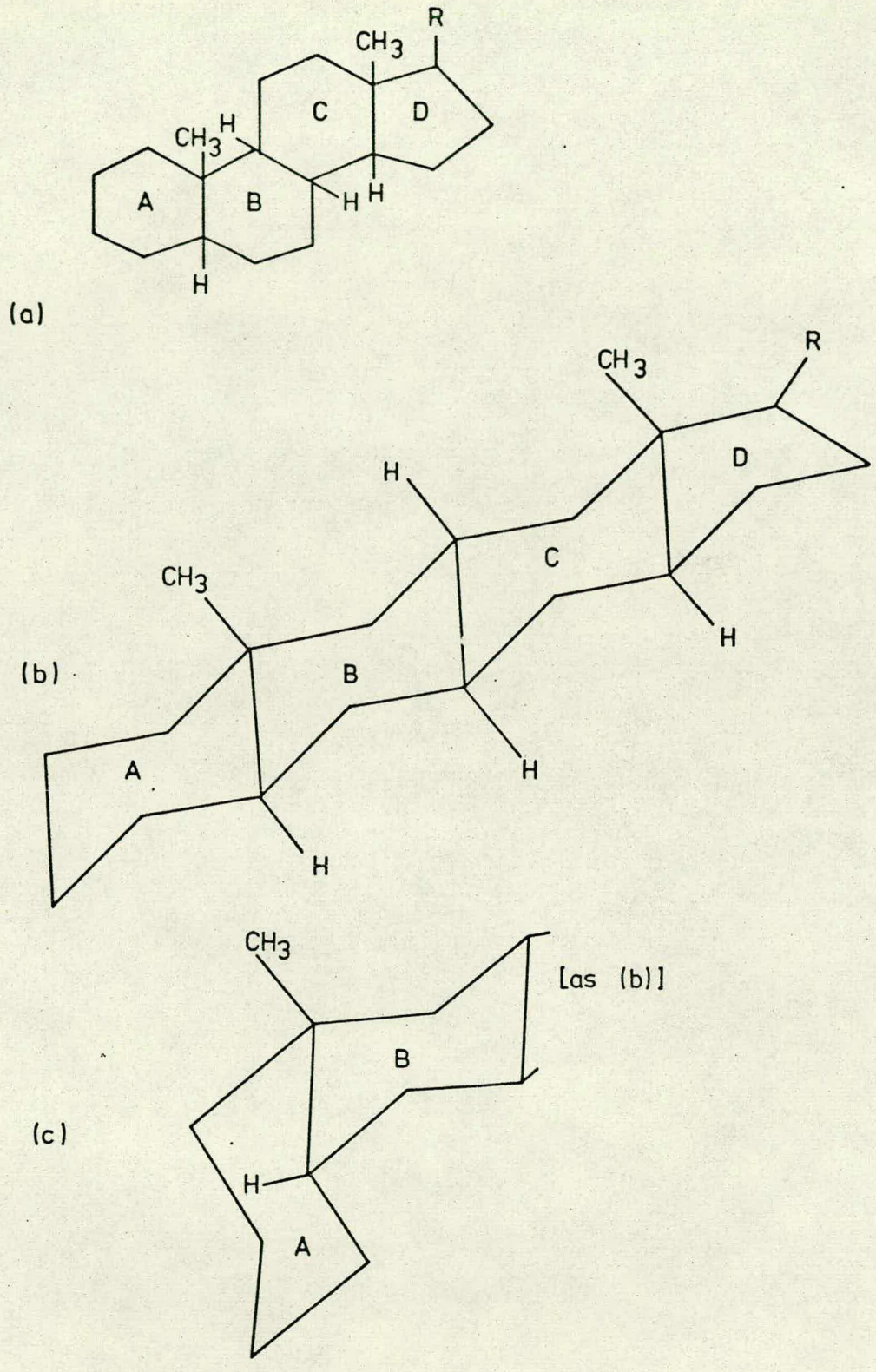


Fig. 5. The steroid nucleus considered by Barton: (a) constitution, (b) and (c) proposed conformations.

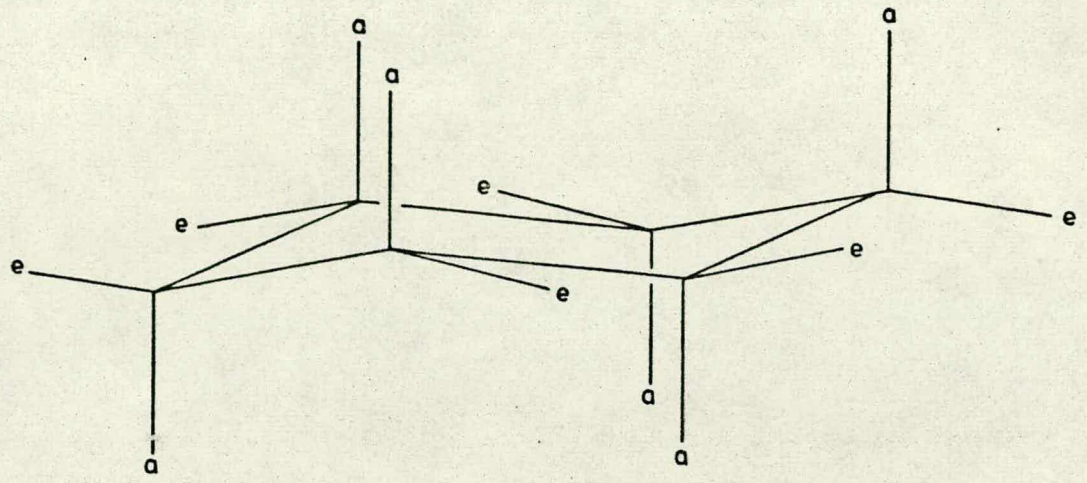


Fig. 6. Equatorial (e) and axial (a) side groups on a cyclohexane ring. The equatorial groups are approximately in the 'plane' of the ring.

Many workers have subsequently continued the investigation of the conformations of cyclohexane derivatives by thermodynamic studies. Eliel, in one of a long series of papers [6], has recently [7] investigated very thoroughly the conformational enthalpies and entropies of hydroxyl groups in various solvents, mainly by NMR studies on substituted cyclohexanols, a study which is of particular relevance to problems to be discussed later in this thesis.

Early work on carbohydrates was carried out by Reeves [1,8-10], who comprehensively studied ring conformations and established the relative stabilities of the different pyranose ring conformations. Much of his work was carried out by the investigation of cuprammonium complexes.

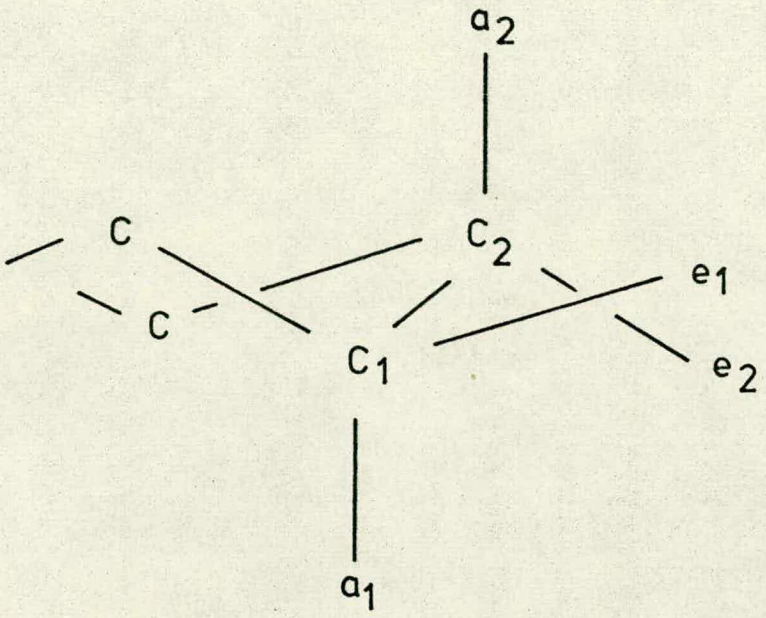


Fig. 7. In the 'ideal' ring the distances e_1e_2, e_1a_2 and e_2a_1 are all around 2.9 Å, whereas a_1a_2 is much greater: about 3.7 Å.

'Cuprammonium' forms complexes with pairs of hydroxyl groups where the distance between the oxygen atoms is around 2.5 Å, thus, (fig. 7) in the chair form, two consecutive equatorial hydroxyls, or one equatorial and one axial will complex, whereas two axial will not. Since different conformations have different numbers of such complexing pairs, the degree of complexing gives valuable conformational information. Reeves followed these complexing reactions by optical rotation and conductimetric measurements and determined and confirmed structures for both monosaccharides. and some polysaccharides. He also used his data to demonstrate [11] the relative stabilities of chair and boat forms of the

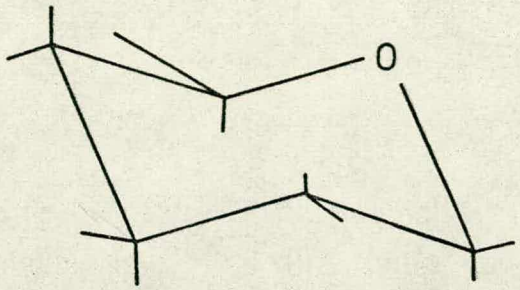


Fig. 8. Reeves C1 ring conformation.

pyranose ring and thus that nearly all monosaccharides exist in the C1 conformation (fig. 8).

Much of the thermodynamic work on monosaccharides has been systematised by Angyal and his co-workers. Angyal [12,13] has developed an empirical method for calculating free energies of monosaccharides, based on thermodynamic data, which gives good agreement with experimental data, particularly that concerning ring conformation and anomeric equilibria. This method is detailed in section 3.2.

The literature on the ring conformation of monosaccharides and their derivatives as established by experimental

techniques is considerable and has been extensively reviewed, most recently by Durette and Horton [14].

1.5 EXPERIMENT-BASED WORK - X-RAY DIFFRACTION

Traditional methods of x-ray crystallography on crystals of mono- and oligosaccharides have yielded much structural information including direct evidence of ring conformations. Work in this area has recently (1972) been reviewed, *inter alia*, by Durette and Horton [14], and an analysis of several such determinations has been published by Arnott and Scott[15].

However, it is as a tool for the direct determination of linkage conformation in polysaccharides [16] in the solid state that x-ray diffraction yields its most interesting conformational information.

If the linkage conformation is regular, then the polymer chain will describe a helix, and further, if such a helix can be stabilised by, for example, hydrogen bonding between its turns, then the regular conformation will be favoured. Alternatively, in different geometrical circumstances, several helices, commonly two or three, may intertwine and be similarly stabilised by hydrogen bonding between the strands.

Most polysaccharides do not crystallise, but many can nevertheless be pulled into fibres in which helically ordered sections or 'crystallites' are oriented more or less

along the fibre axis. X-ray diffraction photographs then yield information about the structure along this axis (fig.9), namely the size of the repeat unit, h/i, where h is the pitch of each of i intertwined helices, and n, the number of monomer units per turn, the former from the layer line spacing and the latter from the distribution of reflections of meridional intensities. From n and h and the geometry of the repeating monosaccharide residue may be obtained the few possible pairs of ψ/ϕ values, (see section 4.2). Most of these can usually be eliminated because of impossible steric clashes (see section 1.10), and with the help of models or computer searches a unique structure may often be assigned. This or a similar approach has had many triumphs outside the polysaccharide field, notably the elucidation of the DNA structure by Watson and Crick [17,18] from Wilkins' x-ray data in 1953 and the postulation of the hydrogen bond stabilised α-helix in proteins by Pauling, Corey and Branson in 1951 [19].

In the polysaccharide field, Rees [20,21] has classified the known and postulated chain shapes for homopolysaccharides. For two of his four generic types (fig.10) he has correlated hard-sphere calculations with available x-ray data.

Of Rees type A polysaccharides, the most important is cellulose (I), on which structure determinations by physical methods have been many, [22] with x-ray work dating as far back as 1913. Although various conformational schemes

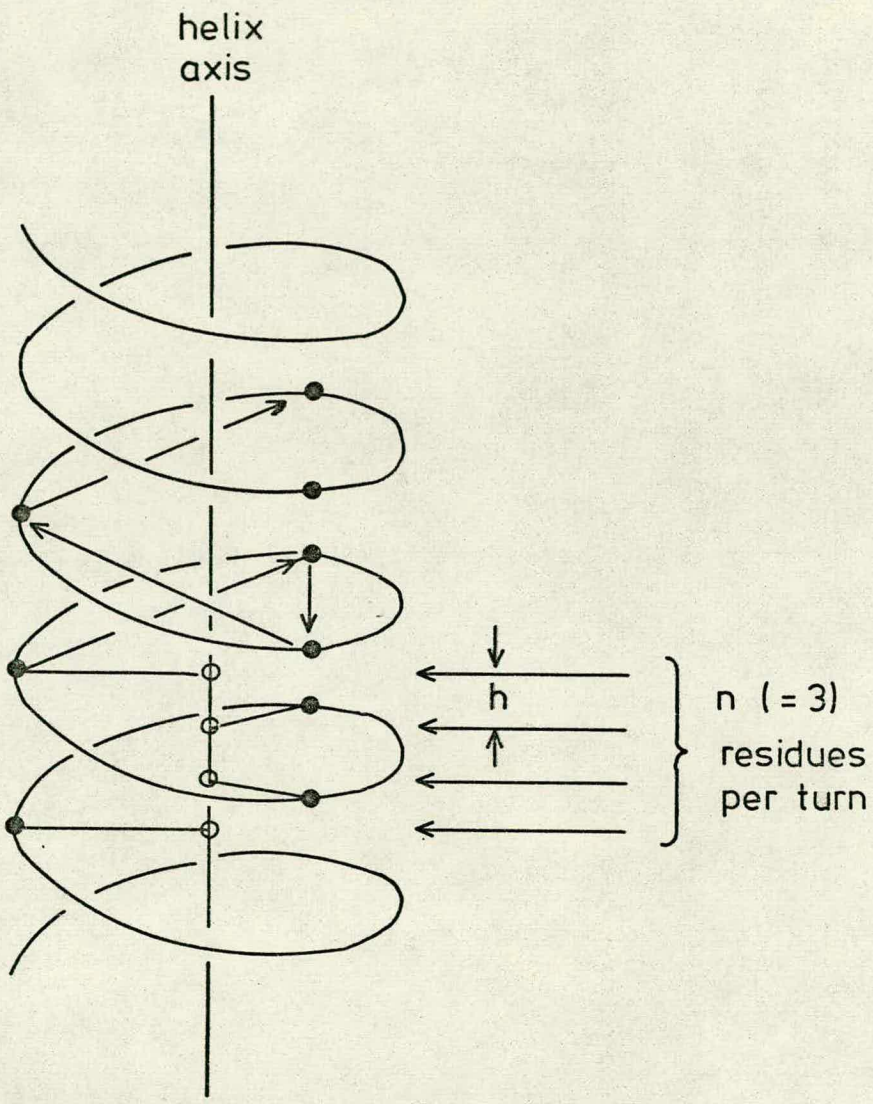


Fig. 9. Helix parameters n and h. The black dots represent equivalent atoms (such as glycosidic oxygens) on successive residues of a regular homopolysaccharide.

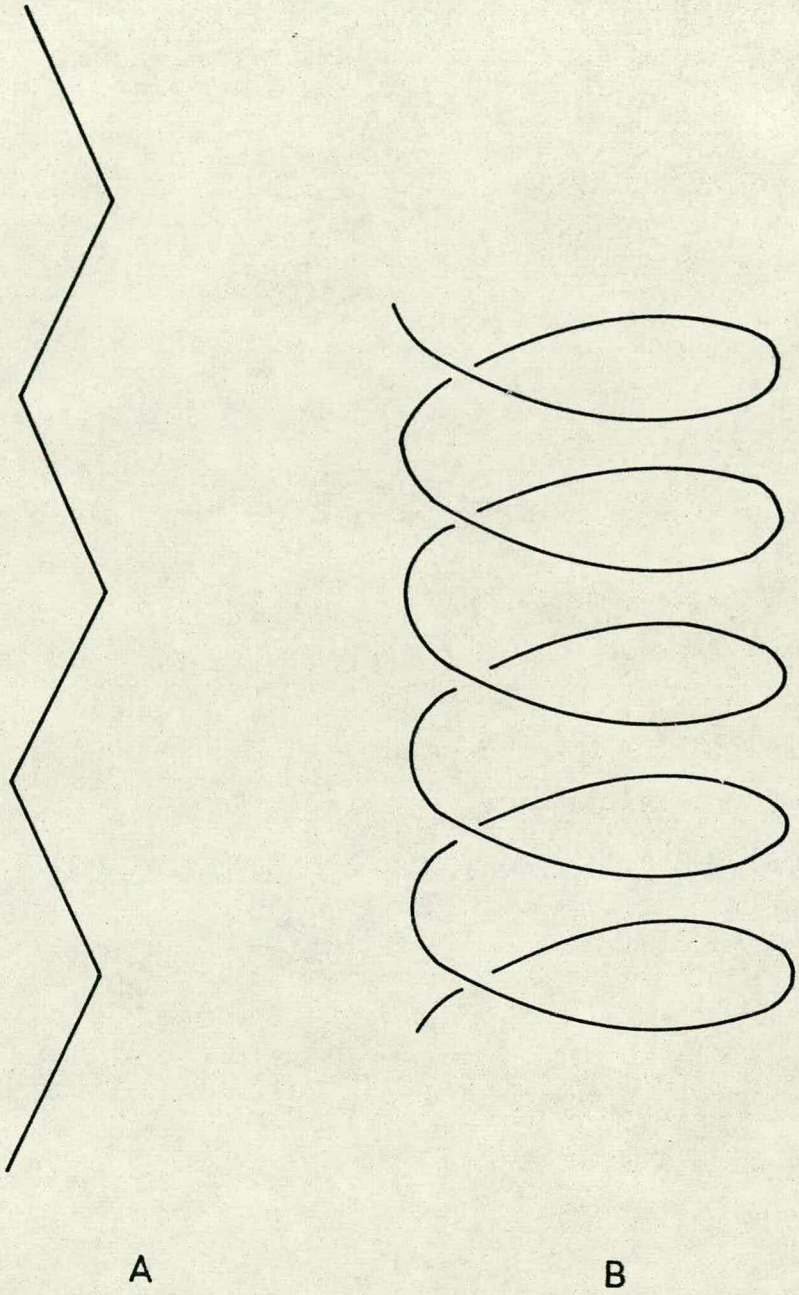
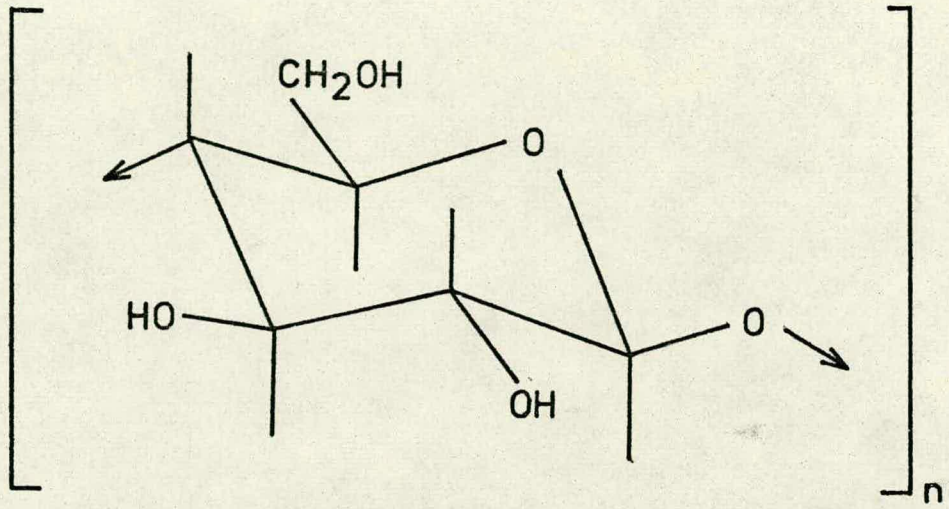


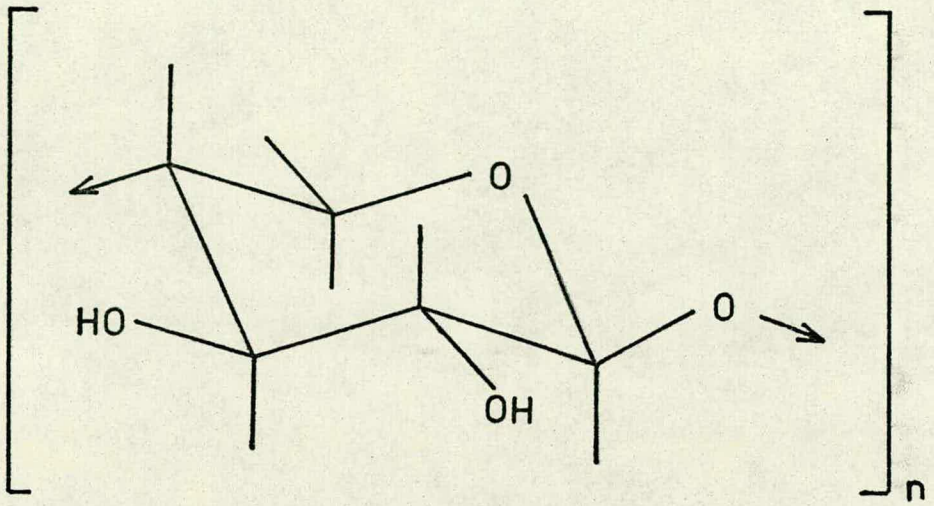
Fig. 10. Backbone shapes of two of Rees' generic types of polysaccharides: A, the flat ribbon with $2 < \pm n < 4$ and h large; B, the spring with $2 < \pm n < 10$ and h small.



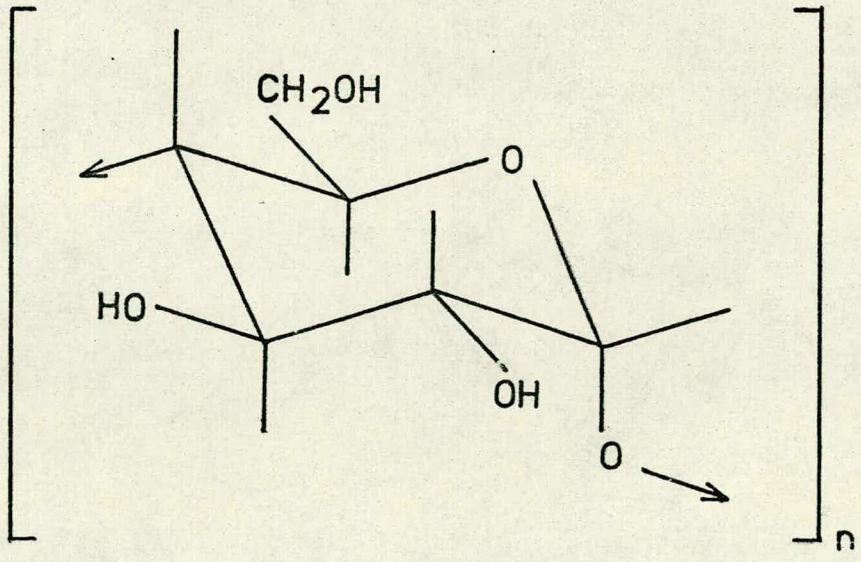
I. Cellulose: β-1,4-glucan.

have been postulated, the detailed structure remains the subject of debate. Notable amongst the successes of the combined x-ray/calculation approach is β-1,4 xylan (II) which was studied by Settineri and Marchessault [23] who showed it to be a three-fold left-handed type A helix. Several other 1,4 diequatorially linked polysaccharides and some 1,4 diaxially linked ones have been shown to adopt the type A structure [21].

Amylose (III) and its derivatives show a wide diversity of forms. In contrast to the open ribbon-like structure of cellulose, amylose, when complexed with iodine or a wide range of other substances, forms a tight (i.e. low h) type B helix known as its 'V form' which has a six-fold (with iodine, or rather



II. β-1,4-xylan



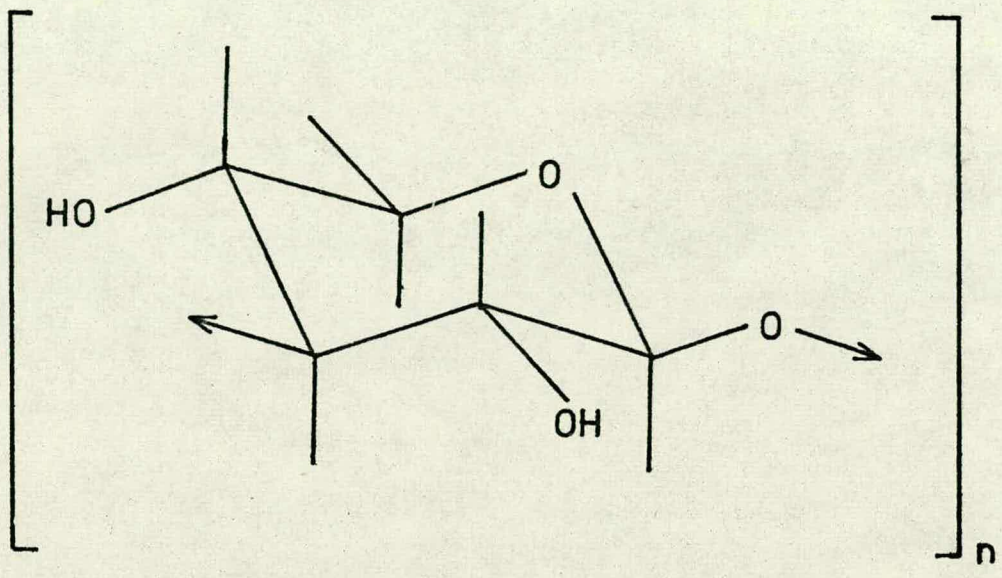
III. Amylose: α-1,4-glucan

more with larger complexing agents) helix, thought to be left handed [24]. The complexing agent is apparently accommodated in the 'hole' along the helix axis.

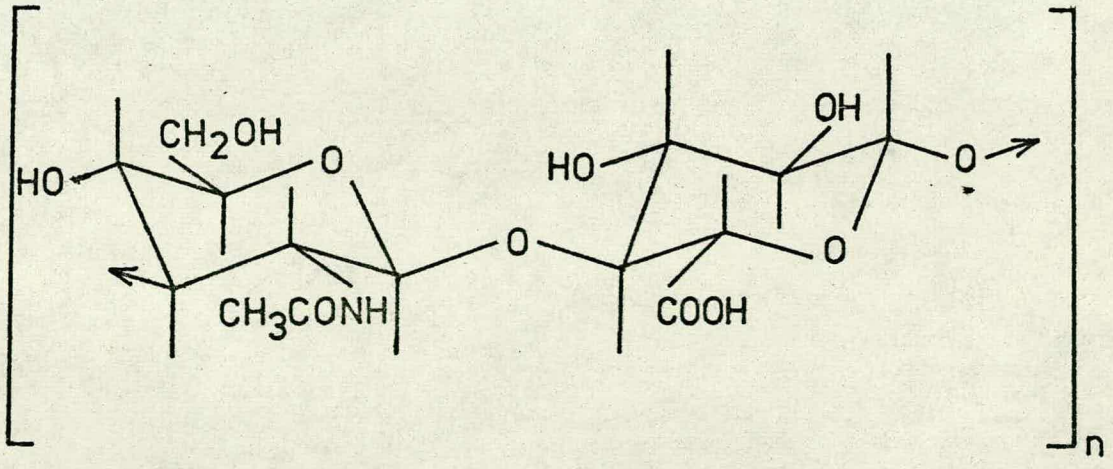
The other regular uncomplexed form of amylose, the 'B form' is clearly a more open structure, since the value of h is 10.4 Å against 7.9 in the V form. Blackwell, Sarko and Marchessault [25] have shown that this could be a hydrogen bond stabilised single helix: alternatively, a double helix with a pitch of 21 Å has been suggested [26]. This latter structure has certain advantages but is difficult to reconcile with some hard sphere and energy calculations[21]. Other derivatives of amylose show an even further elongated chain.

Another well characterised [27] type B polysaccharide is β-1,3 xylan (IV) which exists as a six fold triple helix with a projected residue height of around 3.0 Å. The O(2) atoms on each residue are favourably placed for the hydrogen bonding necessary to stabilise this type of structure. It is thought [21] that β-1,3 glucans will adopt this same structure since the configurational difference on the ring occurs on the outside of the helix, and this is borne out by diffraction studies.

Several alternating polysaccharides have also been studied. Hyaluronic acid (V) has recently been studied by Atkins and co-workers [28] in an acid and a salt form, each of



IV. β-1,3-xylan



V. Hyaluronic acid: alternating copolymer of 2-acetamido-2-deoxy-β-D-glucopyranose and β-D-glucopyranosyluronic acid residues.

which exhibited a different ribbon like structure, whilst Dea et al [29], using samples with which care had been taken to preserve the native state, have shown double helical structures, again for both salt and acid. The closely related κ- and ι- carrageenans (VI) have been studied in detail [30], and have been shown to exist as double helices with three disaccharide units per turn, this being the only interpretation of the diffraction pattern consistent with model building calculations.

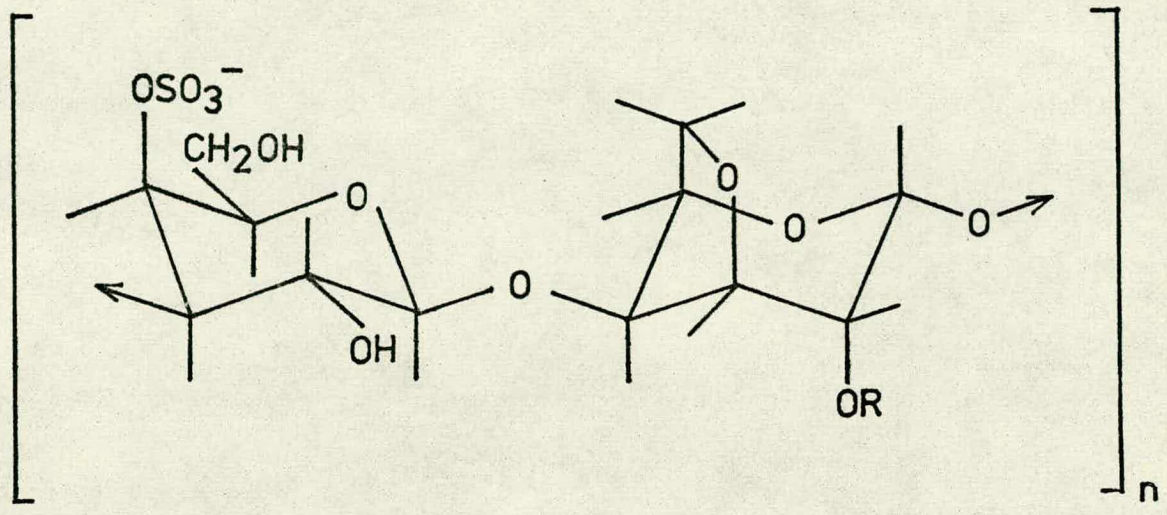
Further treatment of the diffraction pattern yielding refinement of the detailed structure has not been extensively applied to polysaccharides, but a recent paper by Arnott and Scott [15] shows the potential of the technique.

Fibre x-ray diffraction, thus, is a powerful tool which, when taken with relatively simple calculation methods, permits the linkage conformations of many polysaccharides in the solid state to be established.

1.6 EXPERIMENT-BASED WORK - NUCLEAR MAGNETIC RESONANCE

Whilst x-ray diffraction gives information on the solid state structure of molecules, NMR finds application in solution in both non-aqueous solvents and in water.

For monosaccharide work, Lemieux et al [31] have established that equatorial protons generally resonate at lower fields than axial and have thus formulated [32,33]



VI. Carrageenans:

(i) κ-carrageenan: R=H, alternating copolymer of β-D-galactopyranose 4-sulphate and 3,6 anhydro-α-D-galactopyranose residues. In the native state the anhydride residue is partly replaced by its 6-sulphate, galactose 6-sulphate and 2,6-disulphate.

(ii) ι-carrageenan: $R = SO_3^-$, alternating copolymer of β-D-galactopyranose 4-sulphate and 3,6-anhydro-α-D-galactopyranose 2-sulphate residues. In the native state the anhydride residue is partly replaced by galactose 2,6-disulphate.

empirical rules enabling the configuration and conformation of many sugars to be assigned [32-34]. In particular, it has been possible to investigate [12,32,35,36] the position of anomeric equilibria in D_2O solution, and the relative abundance of furanose and pyranose ring forms.

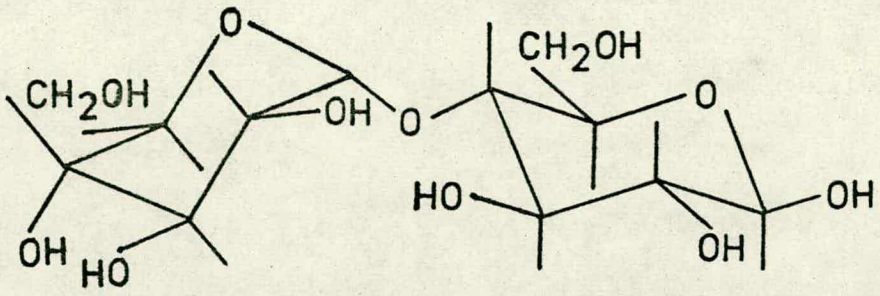
Spin decoupling of identified protons has enabled assignment to be made of ^{13}C signals, and the effect of different axial and equatorial substituents on the ring carbon atoms has thus been studied [37], and would seem to be correlated with conformational stability.

Coupling constants, particularly between hydrogen atoms on adjacent carbon atoms, have been used [38] to obtain information on configurational and conformational properties by the application of Karplus' rule [39] relating the dihedral angle H-C-C-H to the coupling constant.

Linkage conformations of di- and polysaccharides in solution are of course more variable than in the solid state. If, however, the conformation angles vary only within a restricted range, average ϕ and ψ values may be measurable. If the conformational energy due to changes in linkage conformation is $E(\phi,\psi)$ then the observed $\langle\phi\rangle$ is given as a thermodynamic average by

$$\langle\phi\rangle = \int \phi \exp(-E(\phi)/RT)d\phi / \int \exp(-E/RT)d\phi \qquad (1)$$

and similarly for $\langle\psi\rangle$



VII. β-maltose:
4-O-α-D-glucopyranosyl β-D-glucopyranose.

With this limitation, however, linkage conformations in dimethyl sulphoxide (DMSO) solution have been investigated [40] by proton magnetic resonance involving hydroxy hydrogens, whose resonances show downfield shifts when they are involved in hydrogen bonds. In this way, Casu and co-workers [40] have shown the existence of differing proportions of an O(2) .... O(3)' hydrogen bond in maltose (VII) and related compounds.

In other work, Casu and others [41,42] have shown a relationship between the chemical shift of H(1) and the expected value of ϕ, which may be of value in aqueous systems.

^{13}C work with disaccharides [43] has shown that in

maltose, but not cellobiose or lactose, the spectrum does not derive simply from that of the monomers, and thus that steric effects related to the conformation must be present. These have yet to be analysed.

NMR studies thus give good evidence for certain aspects of conformation in solution, and may readily be combined with other techniques such as optical rotation and infra-red spectroscopy.

1.7 EXPERIMENT-BASED WORK - CHIROPTICAL METHODS

The chiroptical methods of single wavelength optical rotation, optical rotatory dispersion (ORD) and circular dichroism (CD) are applicable to solutions of carbohydrates in certain solvents, notably water.

Single wavelength, usually the sodium D-line, studies have found application in several areas. For monosaccharides, empirical rules for the prediction from ring conformation and sidegroup configuration of the optical rotation have been derived by several workers, notably Hudson [44] whose 'rules of isorotation' concern anomers of pyranose derivatives. These rules break down in certain circumstances and were considered by Kauzmann [45], who derived from quantum mechanical considerations a treatment giving the total rotation as a sum of all two-atom, three-atom and so on interactions. Whiffen [46] and later Brewster [47] put this on a quantitative basis by establishing the magnitudes of these

interaction terms, and this predicts values agreeing well with experiment.

For polysaccharides, Rees [48,49] has introduced the concept of 'linkage rotation' which he defines as the difference between the rotation due to two monosaccharides and that observed for a disaccharide. By application of Brewster's rules he has shown the relationship between linkage rotation $[\Lambda]$ and the conformational angles ϕ and ψ (fig. 11) to be

$$[\Lambda] = 105s-120\ (\sin \Delta\phi + \sin \Delta\psi) \qquad (1)$$

where $s = \pm 1$ for β or α linkage, and $\Delta\phi$ and $\Delta\psi$ are the ϕ and ψ differences from the positions of eclipse of the C-H and O-C bonds (fig, 11). This rule provides a very useful basis for the comparison of calculated and experimental linkage conformations. (See section 4.3).

A further application in the polysaccharide field is provided by the use of D-line rotation to follow the coil to double helix transformation in, for example, ι- and κ-carrageenan [50] (VI). In a similar manner cooperative stabilisation of polysaccharide helices by galactomannans has been shown to occur.[51]

ORD and CD give identical information in principle, although in practice, CD spectra are easier to interpret. This is because an ORD signal tails off slowly (fig. 12) to either side, thus interfering with adjacent ones, whereas the CD spectrum consists of reasonably narrow Gaussian peaks.

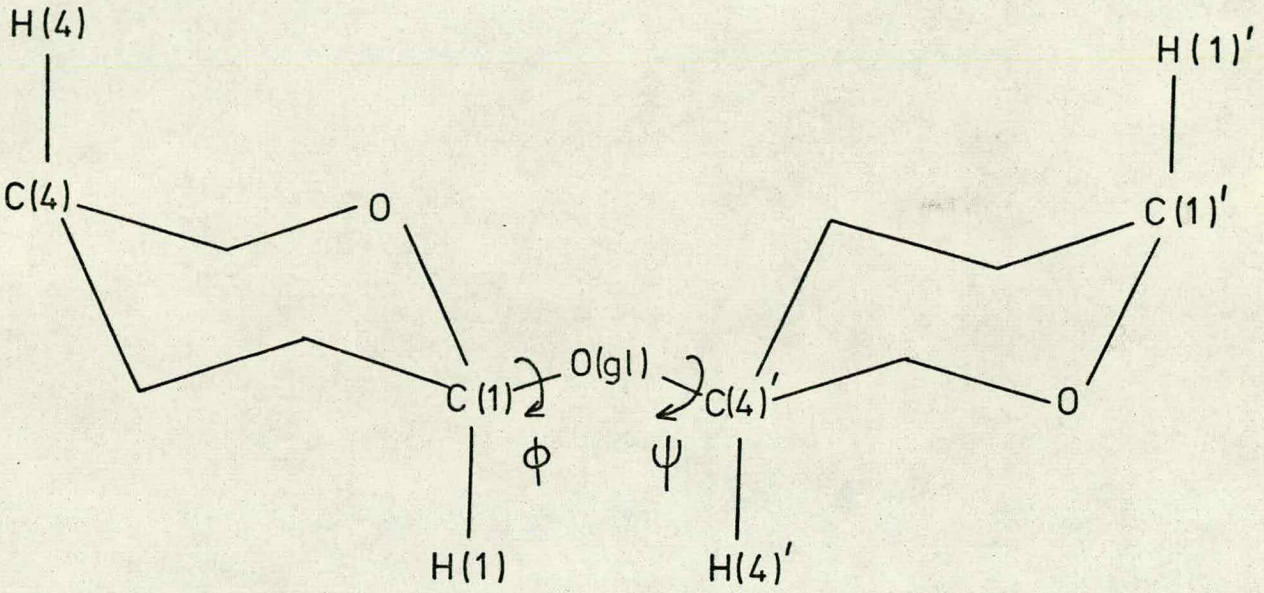


Fig. 11. Zeroes of ϕ and ψ.
$\phi = \psi = 0$ when C(4), C(1), O(gl), C(4)', C(1)' are coplanar, and when C(4) and C(1)' are closest.
$\Delta\phi = \Delta\psi = 0$ when H(1), C(1), O(gl), C(4)', H(4)' are coplanar, and when H(1) and H(4)' are closest.

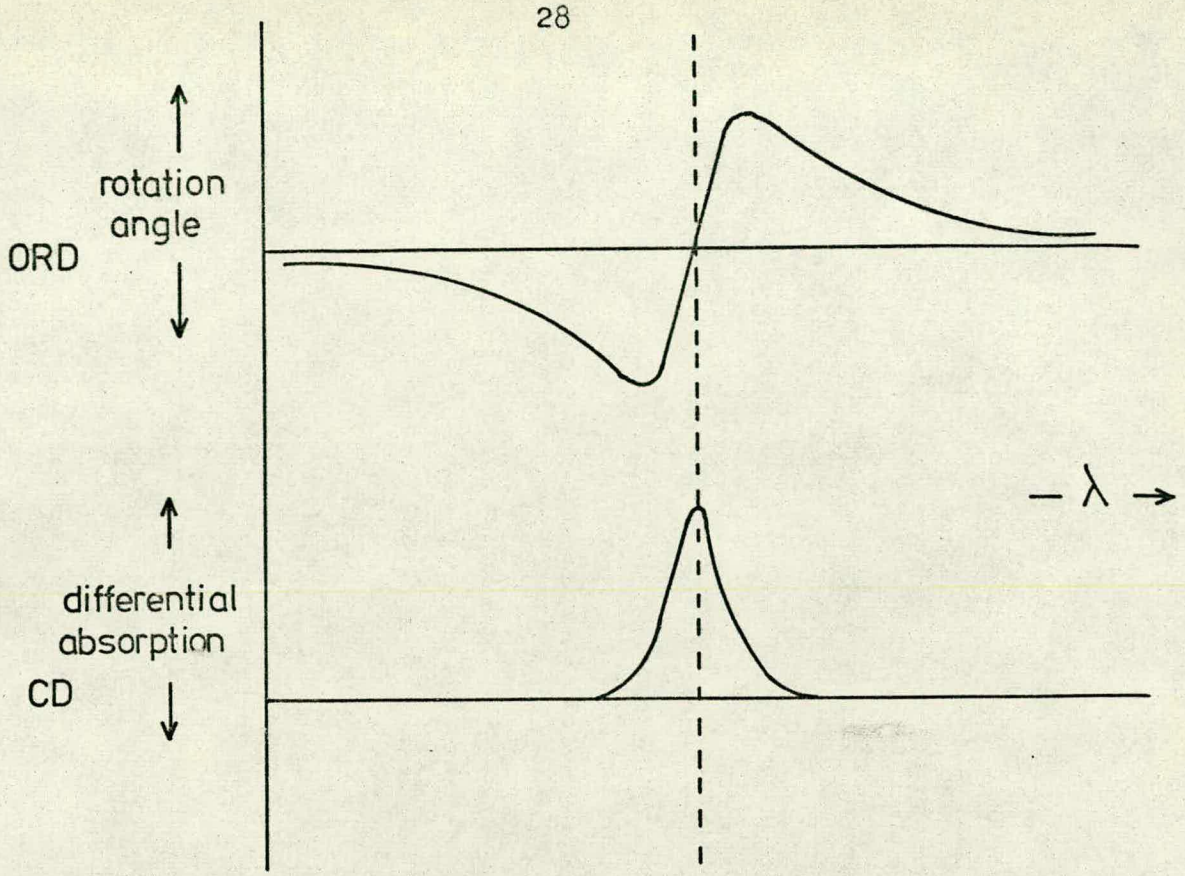


Fig. 12. ORD and CD spectra from the same chromophore.

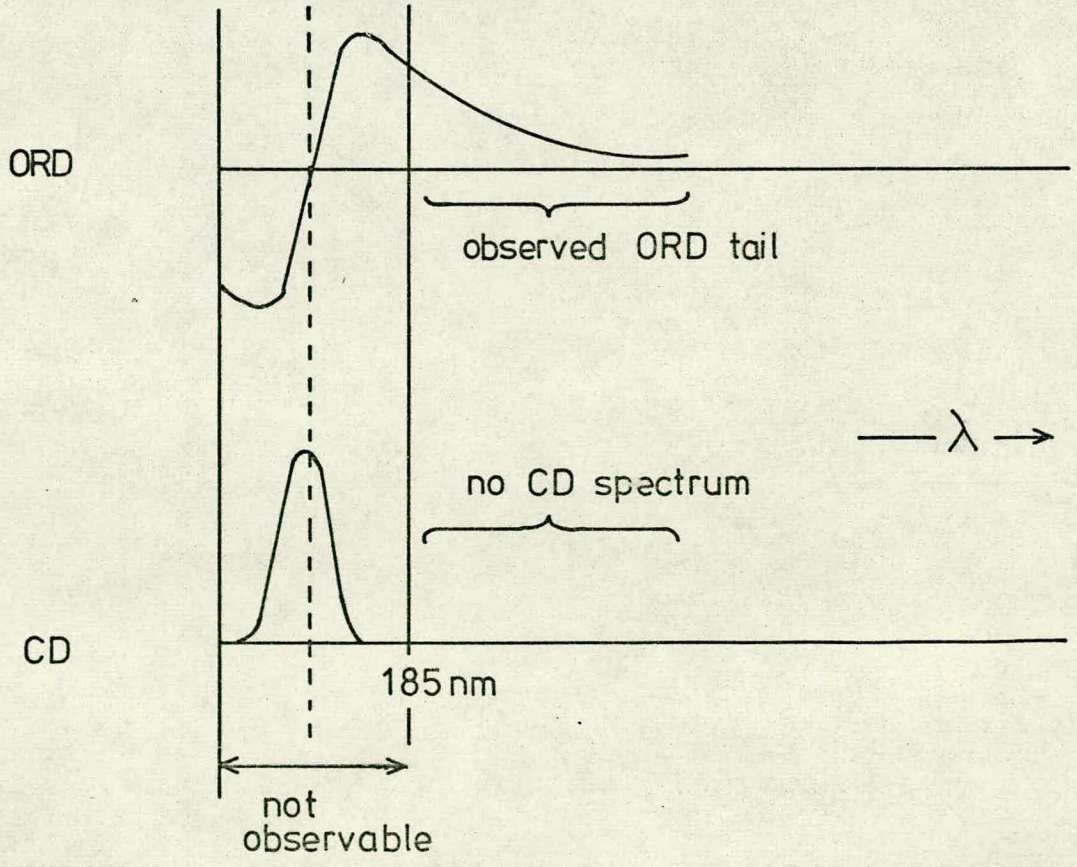
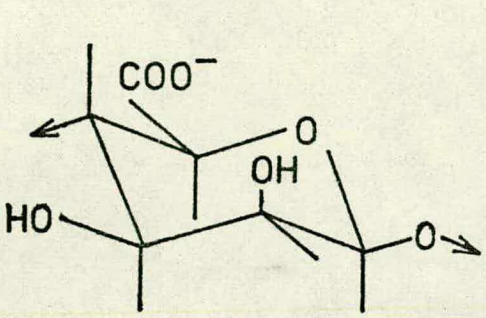


Fig. 13. A just hidden CD peak reveals itself in the ORD spectrum.

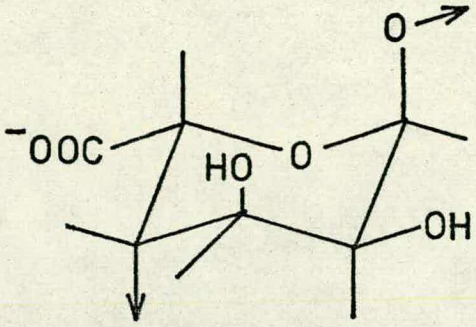
Because, however, the high-frequency limit of present UV sources (about $\lambda = 185$ nm) renders part of the spectrum inaccessible, it is occasionally possible to analyse the observed ORD tail (fig. 13) of an absorption just outside the observable region.

Electronic transitions causing differential UV absorption are termed chromophores and the most common one studied is the carbonyl $n \rightarrow \pi^*$ transition. Empirical rules have been derived for this and other common chromophores, relating the spatial distribution of atoms close to the chromophore with the sign of the CD peak. In carbohydrates however, the carbonyl group is not common, and use must be made of chromophores such as carboxyl or amide groups, either naturally occurring, or suitably added.

Of those polysaccharide derivatives which occur naturally with chromophores, uronic acids have been studied by Listowsky [52] et al and by Morris [53] and co-workers. Thom [54] has studied the gelation of alginates (VIII) by diffusion of calcium ions into sodium alginate solution and has used CD to follow the conformational changes involved. Similarly, Grant [55] has studied pectin (IX) which in its low methoxy (30% methyl esterification) form behaves similarly to alginate with calcium ions. In its high (70%) methoxy form it gels only in the presence of additives which reduce the water activity, forming a non-melting gel in the presence of sucrose, and gelling reversibly with temperature with

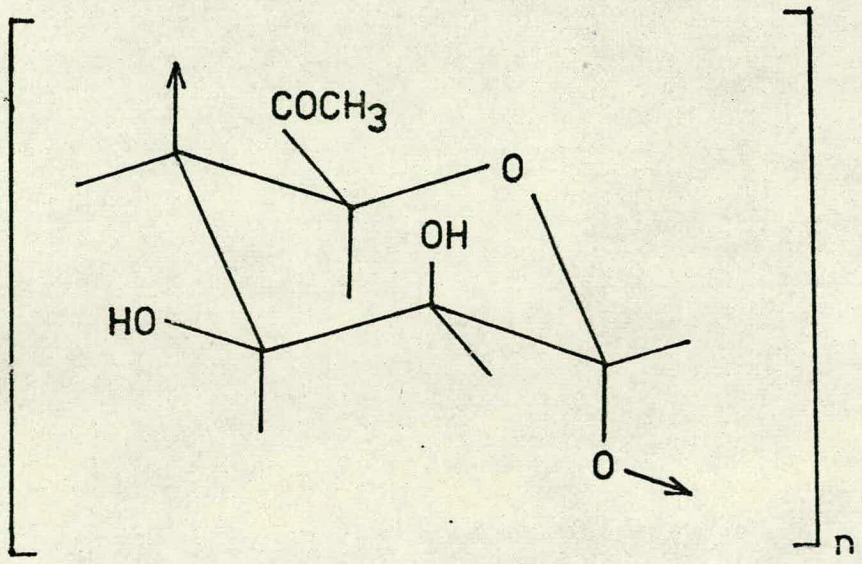


(i)



(ii)

VIII. Alginate. Alginic acid is a copolymer of β-D-mannopyranosyluronic (i) and α-L-gulopyranosyluronic (ii) acid residues. The structure is not, however, regularly alternating.



IX. Pectin: polymer of methyl α-D-galactopyranuronate.

ethylene glycol. Many other examples of this approach exist.

For the majority of polysaccharides, which do not possess accessible chromophores, recourse must be made to the substitution for suitable hydroxyl groups of such derivatives as acetates, but the effect which this has on the conformation is, of course, not apparent. Alternatively, anionic polysaccharides (sulphates, carboxylates, etc.) can combine [56] with certain cationic dyes (e.g. acridine orange, methylene blue) giving complexes whose spectra differ from those of the free dyes in a way which may be related to the polysaccharide structure.

The use of CD in polysaccharide work has recently (1972) been reviewed by Morris and Sanderson [57].

1.8 EXPERIMENT-BASED WORK - OTHER TECHNIQUES

Several other physical methods have been used for the determination of carbohydrate conformation and configuration in particular systems.

Infra-red spectroscopy has been used, particularly for the determination [58,59] of configuration of the anomeric carbon atom (C(1)). Although this may be more readily attainable by NMR in solution, the infra-red method is also applicable to the solid state.

Some evidence for intramolecular hydrogen bonding has been obtained [60,61] by infra-red spectroscopy of very

low concentration carbon tetrachloride solutions, on the basis of the lowering of the O-H bending frequency in hydroxyl groups involved in hydrogen bonding.

Dipole moment measurements have also certain limited application.

Stoddart [62] has recently (1971) reviewed physical methods of configuration and conformation determination.

1.9 CALCULATION METHODS - PRINCIPLES

The problem of calculating conformational energies has been tackled in various ways. Each starts by assuming certain constraints to the geometry of the molecule based on experimental evidence. The subsequent treatment may be classified into four broad types:

i) simple treatments based on observed energy differences between different conformations or configurations;

ii) the 'hard-sphere' approach where atoms are assumed to be hard and to rotate freely about single bonds;

iii) the potential function approach in which the atoms are assumed to interact classically according to postulated potential functions; and

iv) the quantum mechanical approach where molecular orbitals are calculated for a given conformation.

For carbohydrates, it is types (ii) and (iii) which have achieved the greatest success. Type (i) approaches cannot usually deal with linkage conformations: type (iv) calculations are usually impracticably long or must be

drastically simplified.

Williams et al [63] have reviewed (1968) calculation methods for all types of molecule.

1.10 CALCULATION METHODS - HARD SPHERE

A great deal of information has in the past been obtained by the use of ball-and-stick or space-filling molecular models. In particular this applies to 'crowded' molecules, where what would otherwise be free rotation about single bonds is hindered by clashes between pendant groups.

The hard-sphere calculation approach mathematically models this situation and permits a much more effective and efficient survey of such problems.

The examination of linkage conformations in disaccharides, and by extrapolation in polysaccharides, lends itself well to this technique. The necessary prerequisite information comprises

a) cartesian coordinates of all the atoms in the sugar unit whose positions are fixed by assuming fixed ring geometry, and

b) radii of the atoms involved.

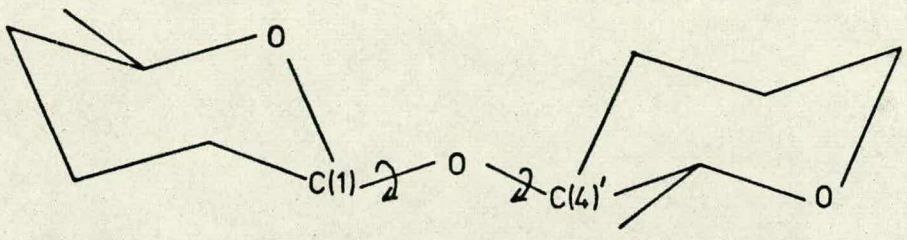
In practice the coordinates are obtained from crystal structure determinations, possibly modified, and the radii are assumed to be the van der Waals atomic radii [64].

The procedure is then to step ψ and ϕ each through

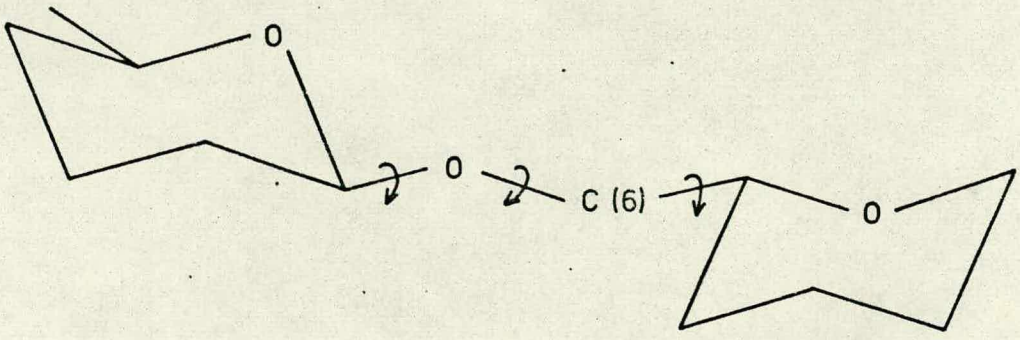
small (e.g. 10°) steps and at each generated conformation to calculate the distances between each atom on one residue with each on the other. As soon as one such distance is found to be less than the sum of the radii of the atoms involved, the conformation is rejected: if none is, then it is accepted. Typically, for 10° steps, with $36^2 = 1296$ conformations only around 0 to 50 are accepted.

A slightly more sophisticated method is to allow for some deformation of the molecule to alleviate slight clashes. This can be accommodated by assuming two radii, the outer being as before and the inner possibly 0.2Å less. Conformations may then be classed as 'fully allowed' as before, or 'allowed with compression' if one or more distance falls between the two limits.

Several workers have used this technique to back up experimental work, particularly x-ray diffraction studies. Settineri and Marchessault [23] have thus studied β-1,4 xylan, and Sarko and Marchessault [65], amylose triacetate. Rees and co-workers [20,30,66] have extensively used hard-sphere methods. For β-glucans linked 1,2, 1,3, and 1,4, they have shown [66] how the conformation predicted may be correlated with optical activity. An analogous treatment [67] for the 1,6 glucan, which has an additional degree of freedom (fig. 14), has also been published. The κ- and ι-carrageenans (VI) have been studied [30] by the combined hard-sphere and x-ray technique, providing the first known example of a polysaccharide double helix. In an



(a)



(b)

Fig. 14. (a) 1,4 linkages (also 1,2 and 1,3) have two degrees of freedom, whereas (b) 1,6 linkages have three.

extended study [20], the chain conformations of a variety of homopolymers of various glucans, galactans, mannans, xylans and arabinans have been determined and classified.

Clearly the hard-sphere method has many successes. Its limitations lie in its inability to predict the effect of long-range forces such as determine largely the pendant group conformation in monosaccharides, and for this a potential function method is necessary.

1.11 CALCULATION METHODS - INTRAMOLECULAR POTENTIAL FUNCTIONS.

The potential function method makes similar assumptions to the hard-sphere method but assumes that for any pair of nonbonded atoms i and j separated by distance r, there is an interaction potential V(r,i,j). This is usually considered as composed of three terms, V_{rep}, a hard repulsion at low r, V_{att}, an attractive term extending to somewhat greater r, and V_{pol} a charge interaction term, often taken as coulombic and varying as r^{-1}: this may be attractive or repulsive. In addition to these two-atom terms, others may be used, such as V_{tor}, a torsional term used to follow more accurately the observed potential barrier on rotation about a single bond (fig. 15; see section 2.4), V_{hb} for hydrogen bonding, V_{as} for angle deformation, V_{bs} for bond stretching or compression, and so on. The total relative conformational potential energy is taken thus as a sum of some or all of these terms.

Many different functions have been suggested or used for $V_{rep} + V_{att}$, mostly based on that of Lennard-Jones:

$$V_{ij} = -Ar_{ij}^{-6} + Br_{ij}^{-n} \qquad (1)$$

where n is generally in the range 8 to 12, or of Buckingham:

$$V_{ij} = -Ar_{ij}^{-6} + C\exp(-\mu_{ij} r_{ij}) \qquad (2)$$

Although the attractive (first) term is theoretically justifiable [68], the repulsive part is empirical and is usually adjusted so that V is a minimum when r is the sum of the van der Waals radii of the interacting atoms.

Brant and Flory [69] have used the Buckingham function and have obtained A and C, assuming $\mu = 4.6$ for all interactions following Scott and Scheraga [70]. In later work, Brant [71] and co-workers have used a Lennard-Jones (LJ) potential with $n = 12$ on maltose. Kitaigorodsky [72] has used the Buckingham potential but effectively with a varying μ. His function is:

$$V_{ij} = -0.14Z_{ij}^{-6} + 3.01 \times 10^4 \exp(-13Z_{ij}) \qquad (3)$$

where $Z_{ij} = r_{ij}/r_o$ and r_o is an empirical equilibrium distance dependent on the types of atoms i and j. Kitaigorodsky derived values for r_o from crystal structure and heat of sublimation data for interactions involving H and C, and Rao and co-workers [73] have extended this to O interactions. Whittington [74, 75] has published work using this function on alginic acid.

Allinger et al [76] have used various similar functions with parameters derived from physical data, notably heats of sublimation, but since their more successful work is on hydrocarbons it is not directly applicable to oxygen containing molecules such as carbohydrates.

Liquori [77] et al have used a more general function

$$V_{ij} = - A\, r_{ij}^{-6} + B \exp(-C\, r_{ij}) r^{-n} \qquad (4)$$

with parameters derived from a variety of physical and theoretical considerations, which reduces to either LJ or Buckingham with suitable choice of A, B, C and n. This has been used successfully for the study of polypeptides.

Ramachandran [78,79] and others have made one of the few available comparisons of energy functions by comparing Brant and Flory and Kitaigorodsky functions on V-form amylose. Although both led to similar conclusions, Skerrett [80] notes that the Kitaigorodsky function follows the hard sphere map more closely.

Skerrett [80] also compares Flory, Kitaigorodsky and Liquori potentials for cellobiose and concludes that although the first two give similar results to the hard sphere calculation, the third does not.

Buckingham type potentials have a disadvantage over Lennard-Jones type in that they take longer to calculate,

a) because allowance must be made for the fact that at low r they unrealistically tend to $-\infty$, and

b) because the exponential takes longer to calculate than the r^{-n} term.

The constants in the LJ potential have been evaluated by Scheraga and co-workers [81-83] and successfully used in polypeptide work. In this work n is taken as 12.

Polar interactions are usually assumed to be well represented by the 'monopole approximation'. In this, 'partial electronic charges' are assigned to each atom and the potential taken as coulombic, viz.

$$V_{ij} = q_i q_j / \epsilon r_{ij} \qquad (5)$$

where ϵ is the effective dielectric constant, to which various values have been assigned, commonly around 3 or 4. The partial charges are usually calculated for non-conjugated systems such as carbohydrates, by the method of del Re [84] (see section 2.3) which is based on an LCAO approach and is designed to match experimentally determined dipole moments.

Torsional potentials have been the subject of much theoretical discussion. For a symmetrical rotation such as in ethane (fig. 15), if θ is the angle of rotation from an eclipsed position, the interaction is usually taken as

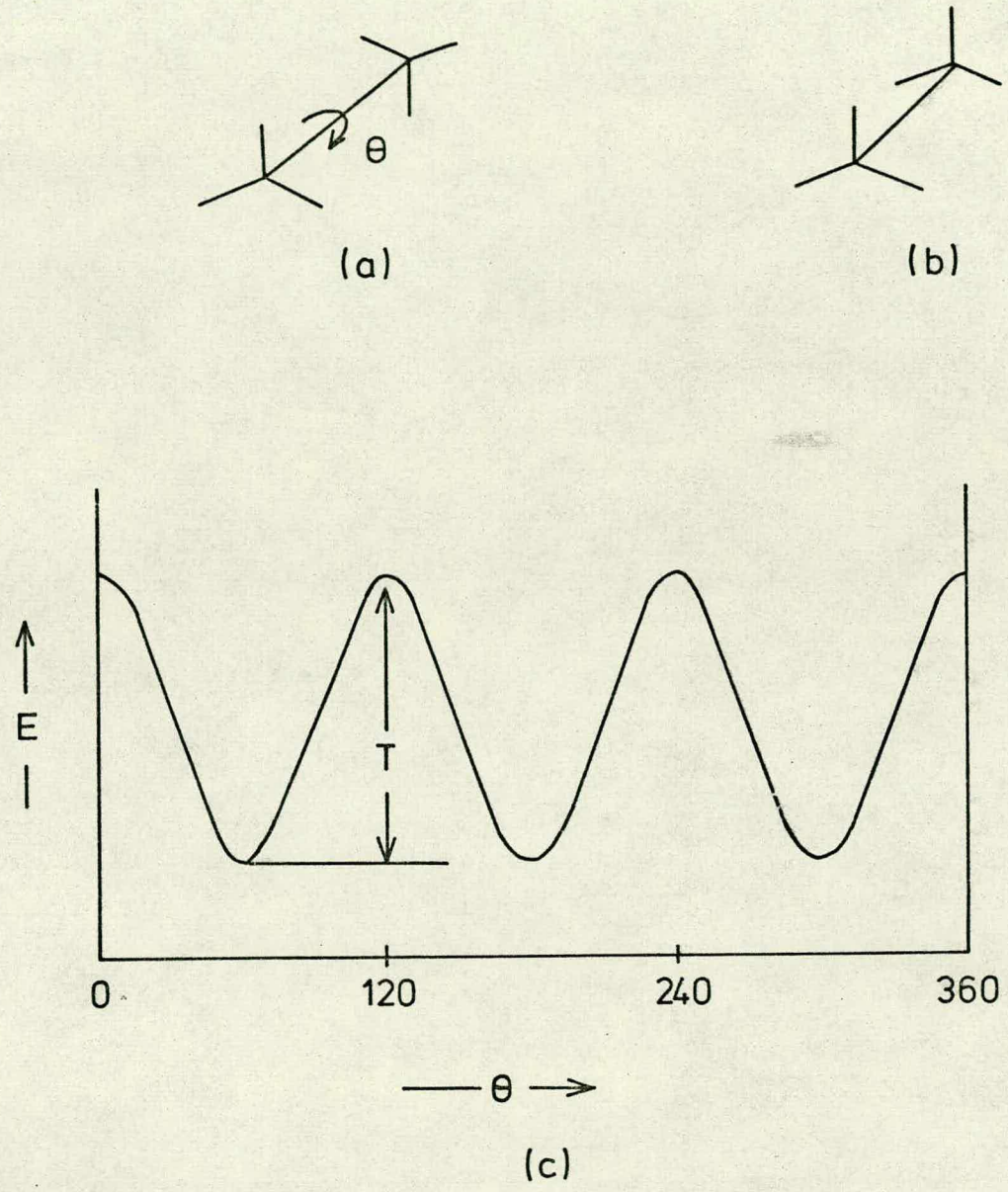


Fig. 15. The torsion barrier in ethane. The staggered conformation (a) is more stable than the eclipsed (b), by energy T which is assumed to vary sinusoidally with θ as in (c).

$$V = \tfrac{1}{2}T(1-\cos 3\theta) \qquad (6)$$

where T is the 'barrier height', estimated from various thermodynamic [85] considerations. Brant and Flory [69], and Scott and Scheraga [82] have both used this approach: Scheraga [86] has reviewed other workers' modifications for various rotations in different systems.

Quantum mechanical treatment [87,88] has shown torsional energy to be merely a consequence of the classical approximation used: nevertheless, it is a useful fiction for the matching of postulated potential functions to observed behaviour.

Hydrogen bond potential functions have been postulated on a semi-empirical basis by Lippincott and Schroeder[89,90] and empirically by, notably, Poland and Scheraga [91], Blackwell et al [25] and Brant and co-workers[71]. Although the first mentioned has been used successfully in certain circumstances [82], it contains in general too many adjustable parameters for use in poorly characterised situations. The Poland and Scheraga function is described further in section 2.5.

Geometrical distortions are generally assumed to be harmonic with force constants estimated from spectroscopic data. They are not commonly taken into account, although Gibson and Scheraga [92] have considered them for

dipeptides, and Rao [93] and co-workers have selectively done so for certain pyranose sugars (see section 3.3).

Reviews of potential functions are many in number, notable being Scheraga's [86] (1968), and that of Williams [63] et al (1968).

It is clear that no universally applicable potential function is available. In the absence of detailed comparative studies the choice of a suitable function for a particular system is still a matter of debate. Results of calculations comparable with experiment are still too few, particularly in carbohydrates, to permit a fair assessment of the available functions.

1.12 CALCULATION METHODS - QUANTUM MECHANICAL

The difficulties of even approximate solution of the Schrödinger equation for the simplest molecules are well known, and thus for larger molecules only grossly simplified models are possible.

A great deal of the published work concerns peptides, as is shown by a recent (1971) survey by Pullman[94]. He and co-workers [95] have published some work on maltose, cellobiose and sucrose using his PCILO approximation which yields generally good results, although they conclude that a procedure for optimising the positions of the hydroxyl groups near the glycosidic linkage is indispensible.

Saran and Govil [96] have performed EHT and CNDO calculations on the various chair conformations of glucose, which yield results similar to those obtained by Reeves.

If the available computing resources increase as rapidly in the next ten years as they have in the last, then techniques such as these will undoubtedly become more important; until such time, however, they are economically impracticable, or must be approximated to such an extent as to render them no better than empirical methods.

1.13 SUMMARY

The preceding sections have scanned briefly the state of current knowledge on methods of determining carbohydrate conformation. They are an attempt to show how calculation methods may supplement experimental techniques and in some cases provide useful extrapolations.

In the following chapters various applications of some calculation methods are described together with a correlation with available experimental evidence.

CHAPTER 2

FORMULATION AND OPERATION OF A POTENTIAL FUNCTION

2.1 INTRODUCTION

In section 1.11, previously used potential functions were briefly reviewed. This chapter describes the selection of a function for further study, and the succeeding chapters describe applications to some problems in carbohydrate conformation.

The basic function to be described has four terms:

i) a van der Waals term describing primarily short-range interactions,

ii) a polar term describing longer-range effects,

iii) a bond torsion term, and

iv) a modification to allow for hydrogen bonding.

The atom numbering scheme used is shown in fig. 16.

2.2 VAN DER WAALS POTENTIAL

For reasons stated earlier, the most rapidly calculated van der Waals potential is one of the Lennard-Jones type:

$$V_{ij} = -A\, r_{ij}^{-6} + B\, r_{ij}^{-n} \qquad (1)$$

The value of n has been variously taken from 8 upwards, most commonly as 12, for which the values of A and B have been evaluated by, amongst others, Scott and Scheraga [82,83]. They obtained A from theory and B by the condition that V be a minimum when r was the sum of the van der Waals radii, plus 0.4 Å.

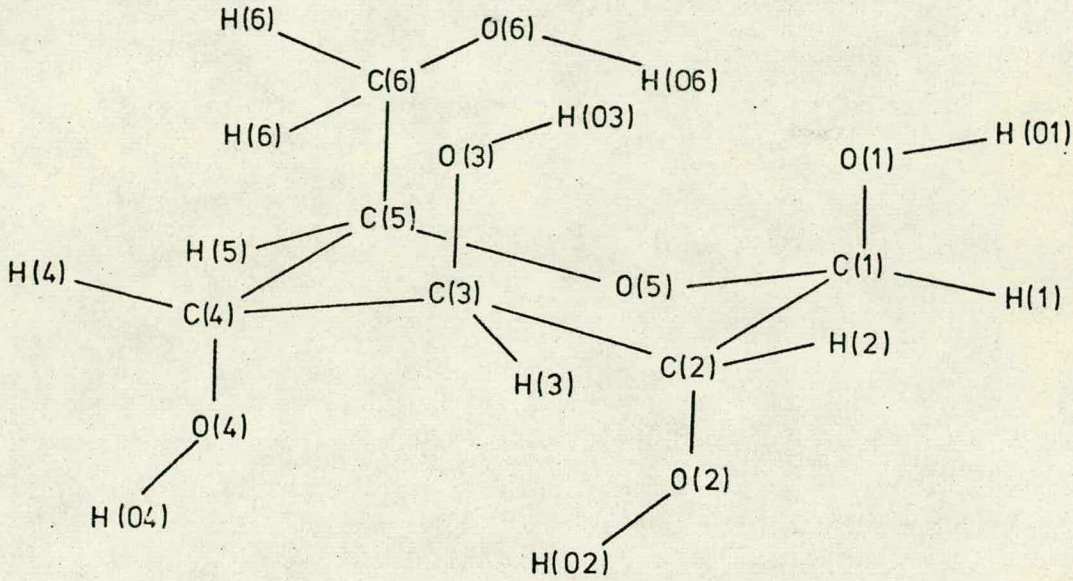


Fig. 16. Numbering of atoms in a sugar. If it is necessary to distinguish each H(6) (or H(5) in a pentose) they are termed H(6a) and H(6b). In a disaccharide, atoms on the reducing and non-reducing residues are termed C(1)R and C(1)N etc.; the glycosidic oxygen may be termed O(gl).

The r^{-12} repulsion is harder (i.e. increases more rapidly) than most exponential-type functions, and for this work, therefore, a value of n = 9 has been used. Some justification of this on the basis of optical rotation studies has been shown by Blann[97].

Using the same assumptions as Scott and Scheraga, if B_9 and B_{12} are the B values for respectively n = 9 and 12, we require

$$dV/dB_9 = 6A\, r_o^{-7} - 9B_9\, r_o^{-10} = 0 \qquad (2)$$

where r_o is the equilibrium distance given by

$$dV/dB_{12} = 6A\, r_o^{-7} - 12B_{12}\, r_o^{-13} = 0 \qquad (3)$$

From (3), $r_o^6 = 2B_{12}/A$

whence from (2), $B_9 = \sqrt{\frac{8}{9}\, AB_{12}}$

The values of A, B_9 and B_{12} are given in Table 1 for all interactions between C, H and O. The shape of the potential is shown in Fig. 17 for the C-C interaction; this shows the typical shallow minimum at r_o followed by a steep rise at lower radii.

2.3 POLAR POTENTIAL

The 'monopole approximation' is a method of assigning fractional electronic charges to each atom in the

Table 1.

Van der Waals potential parameters

Interaction	r_o (Å)	$\bar{A} \times 10^{-2}$ (kcal Å^6 mole^{-1}).	$B_9 \times 10^{-3}$ (kcal Å^9 mole^{-1}).	$B_{12} \times 10^{-5}$ (kcal Å^{12} mole^{-1}).
C-C	3.40	3.70	9.70	2.86
C-O	3.22	3.67	8.18	2.05
C-H	2.90	1.28	2.08	0.38
O-O	3.04	3.67	6.88	1.45
O-H	2.72	1.24	1.66	0.25
H-H	2.40	0.47	0.43	0.045

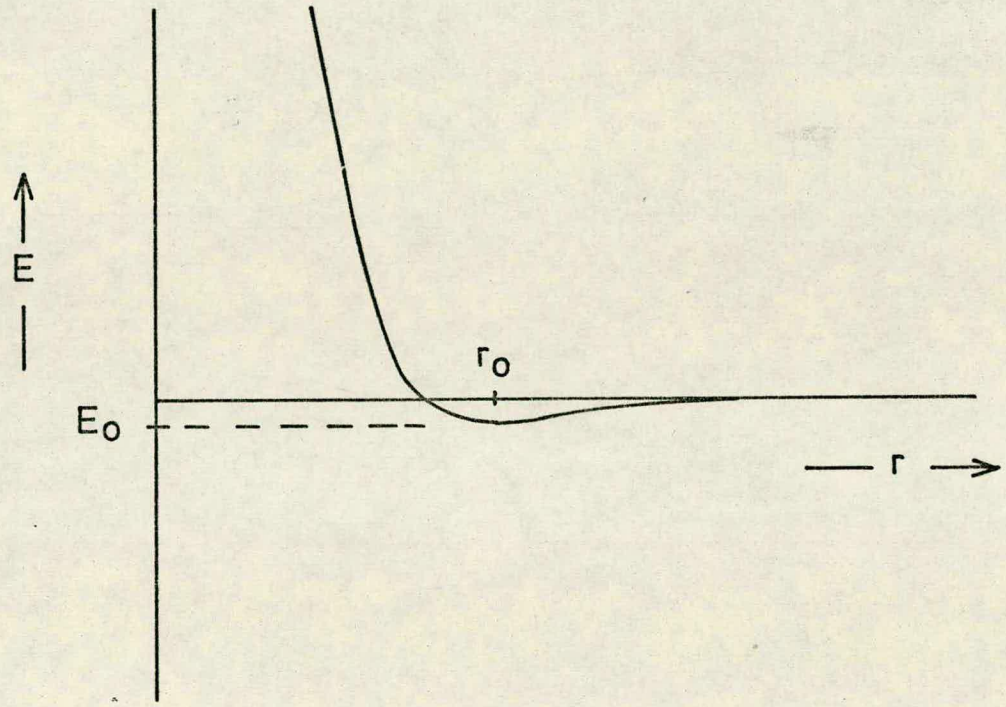


Fig. 17. C-C van der Waals interaction for Lennard-Jones 6-9 potential.

molecule and assuming each to be effectively a spherically symmetrical point charge at the atomic nucleus.

Two methods have commonly been used to assign these charges, one being an attempt directly to reproduce observed bond dipoles in simple molecules [82,98] and the other less directly doing so by a molecular orbital approach whereby the influence of every other atom on the atom considered is taken into account.

This latter method is due to del Re [84,99,100], and is applicable to any fully saturated molecule.

According to him, the 'σ charge' (Q_A) on any atom A may be calculated thus:

If atom A is bonded to atom B, then the contribution to Q_A by bond AB is

$$q_A^{(B)} = \tilde{Q}_A^{(B)} \Big/ \sqrt{1+(\tilde{Q}_A^{(B)})^2} \qquad (1)$$

where $$\tilde{Q}_A^{(B)} = (\delta_B - \delta_A) / 2\epsilon_{AB} \qquad (2)$$

and $$\delta_A = \delta_A^{\circ} + \sum_{\text{all bonded atoms B}} \gamma_{AB}\delta_B \qquad (3)$$

The ϵ_{AB}, γ_{AB} and δ_A° are constants dependent only on the nature of A and B and are given by Scheraga [86]. The charge Q_A is then given by

$$Q_A = \sum_{\text{all bonded atoms B}} q_A^{(B)} \qquad (4)$$

For a molecule containing n atoms, therefore, the method is

i) to solve the n simultaneous linear equations (3) for the δ_A,

ii) to obtain the $\tilde{Q}_A^{(B)}$ from (2) and the $q_A^{(B)}$ from (1), and

iii) to sum these as in (4) to obtain the Q_A.

In practice the linear equations are well behaved and may be solved by most conventional methods [101].

Table 2 gives the σ charges calculated by this method for various carbohydrate molecules: these are in agreement with those which have been published elsewhere, [80,102,103].

With the monopole approximation, the potential due to the charge-charge interactions is taken to be coulombic:

$$V_{ij} = q_i q_j / \epsilon r_{ij}$$

where q_i and q_j are the σ charges as above and ϵ is the apparent dielectric constant.

The value of ϵ has been the subject of much debate. Brant and Flory [104] have selected a value of 3.5 for non-polar solvents, Gibson and Scheraga [105], 3.0 for aqueous work, Sundararajan and Rao [106,107] and co-workers 7.0 and 2.0 for respectively aqueous solution and solid state. All these values were selected as being those that

Table 2

Calculated atomic partial charges

Atom	a	b	c	d
C(1)	0.190	0.190	0.177	0.190
C(2)	0.114	0.114	0.113	0.114
C(3)	0.106	0.106	0.106	0.105
C(4)	0.104	0.105	0.104	0.092
C(5)	0.095	0.035	0.095	0.094
C(6)	0.045		0.045	0.045
O(1)	-0.446	-0.446	-0.264	-0.446
O(2)	-0.457	-0.457	-0.457	-0.457
O(3)	-0.458	-0.458	-0.458	-0.458
O(4)	-0.458	-0.458	-0.458	-0.264
O(5)	-0.262	-0.260	-0.264	-0.262
O(6)	-0.457		-0.457	-0.457
H(1)	0.061	0.061	0.060	0.061
H(2)	0.053	0.053	0.053	0.053
H(3)	0.052	0.052	0.052	0.052
H(4)	0.052	0.052	0.052	0.050
H(5)	0.051	0.052 *	0.051	0.050
H(6)	0.053 *		0.053 *	0.053 *
H(O1)	0.303	0.303		0.303
H(O2)	0.302	0.302	0.302	0.302
H(O3)	0.301	0.301	0.301	0.301
H(O4)	0.301	0.301	0.301	
H(O6)	0.302		0.302	0.302

a - hexopyranose
b - pentopyranose
c - dihexopyranose - non-reducing residue
d - - reducing residue

* - two such atoms

gave the best fit of the results of the calculations with experimental data.

It has been argued [86,108] that ϵ is not independent of r, and that whilst at low r it depends only on the atomic polarisability of the atoms concerned, at larger distances the solvent has a greater influence. Ramachandran and Srinivasan[108] have considered theoretically a system of two interacting bodies within a cavity in a medium of high dielectric constant. Although this is a somewhat artificial situation, they conclude that there is a variation of ϵ with r, but that at the values of r of interest in polymer calculations, the normally used values of ϵ are reasonable.

In practice, however, most workers have not taken account of this variation, but have used a constant ϵ, and this has been continued in the work to be described here.

2.4 TORSIONAL POTENTIAL

Early theories of chemical bonding supposed rotation about a single bond to be entirely free, and indeed no case of isomerism due to restriction of this rotation has ever been found.

Kemp and Pitzer [109], however, have shown from the value of the entropy of ethane that a three-fold energy barrier does exist, although of an energy much less than

thermal. Various attempts have subsequently been made [110] to explain this barrier in classical or semi-classical terms, most of which lead to predictions of much smaller barriers than the experimental work shows.

Not until quantum mechanical treatments were applied did any theoretical approach predict a realistic value. Pauling [111,112] has shown that it is probably the exchange interactions (repulsions) between the electron orbitals in the C-H bonds which cause the barrier to rotation and he estimated a value of about 3 kcal/mole, in accordance with experiment. Such quantum mechanical methods, which have recently (1971) been reviewed by Oosterhoff [113], have led to a better understanding of the nature of torsion effects, but no theoretical method is yet a reliable predictive tool for other than the simplest molecules.

Estimates of the sizes of barriers from physical data [114] have been available for many years. French and Rasmussen [85] in 1946 formulated empirical rules for the prediction of barrier heights for rotation about a CH_3-X bond, which give good agreement with experiment. A more exact early determination of the ethane barrier was made by Kistiakowsky et al[115] who obtained a value of 2.75 kcal/mole: the commonly accepted value nowadays is 2.9 kcal/mole.

The form of the barrier has been almost universally

considered to be sinusoidal, thus if the barrier height is T, and the angle of rotation from the minimum energy (staggered) position, θ, the potential is

$$V = \tfrac{1}{2}T\,(1-\cos 3\theta). \qquad (1)$$

Barrier heights as obtained from physical data must be considered as including the van der Waals repulsion terms between the relevant atoms, and it is thus necessary to reduce slightly the value of T.

In the work to be described, three types of rotation are considered. Each is assumed to be represented by a 'model compound', a small molecule containing an analogous rotation, the barrier height of which has been measured. Table 3 lists these model compounds and the values taken for T. The rotations are:

i) the C-O-C glycosidic bond system, represented by dimethyl ether, CH_3-O-CH_3,

ii) the C-O-H system, represented by methanol, CH_3-O-H, and

iii) the C-C-O system (occurring in hexoses in C(5)-C(6)-O(6)), represented by ethanol, CH_3-CH_2-OH, or ethane CH_3-CH_2-H for which more reliable data is available.

The calculated contribution from the van der Waals potential is also shown.

Table 3

Barrier heights in kcal/mole as measured [116] and as used in calculation (see text)

Model Compound	Barrier height	vdW contribution	T
CH_3-O-CH_3	2.72	0.32	2.40
CH_3-O-H	1.07	0.11	0.96
CH_3-CH_2-H	2.90	0.34	2.56

2.5 HYDROGEN BOND POTENTIAL

As in the case of torsion, a general theoretical treatment reproducing well experimental data is not available for hydrogen bonds. A suitable function must therefore be postulated from empirical arguments.

Various such functions have been used (see section 1.11). In the calculations to be described, the requirement is a function which will yield an easily calculated energy given (fig.18) the distances OH (s), O'H (t), and the angle O'OH (θ). Poland and Scheraga [91] have described such a function for use with a 6-12 Lennard-Jones potential: this has been modified as follows for use with the 6-9 potential used in this work.

Poland and Scheraga consider three facts as established, viz.:

1. there is a known, ideal (i.e. minimum energy) O-O' distance, r_o;
2. this corresponds to a known energy minimum, (which will be negative);
3. at much larger O-O' distances than r_o, there exists an attractive potential proportional to r^{-6}.

They also assume fixed bond lengths and angles and take as essential that the potential merges with the 'normal' one for conditions unfavourable to hydrogen bonding. The 'known' facts are thus summarised as the solid line and dot in fig.19.

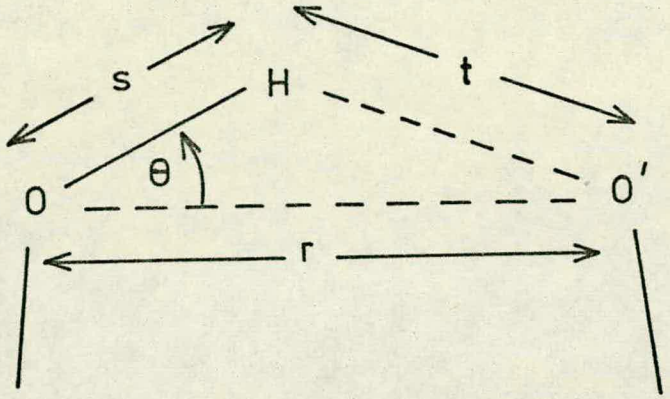


Fig. 18. The hydrogen bond. 'Ideal' parameters are
r = 2.7 Å. s = 1.0 Å. t = 1.7 Å, θ = 0.

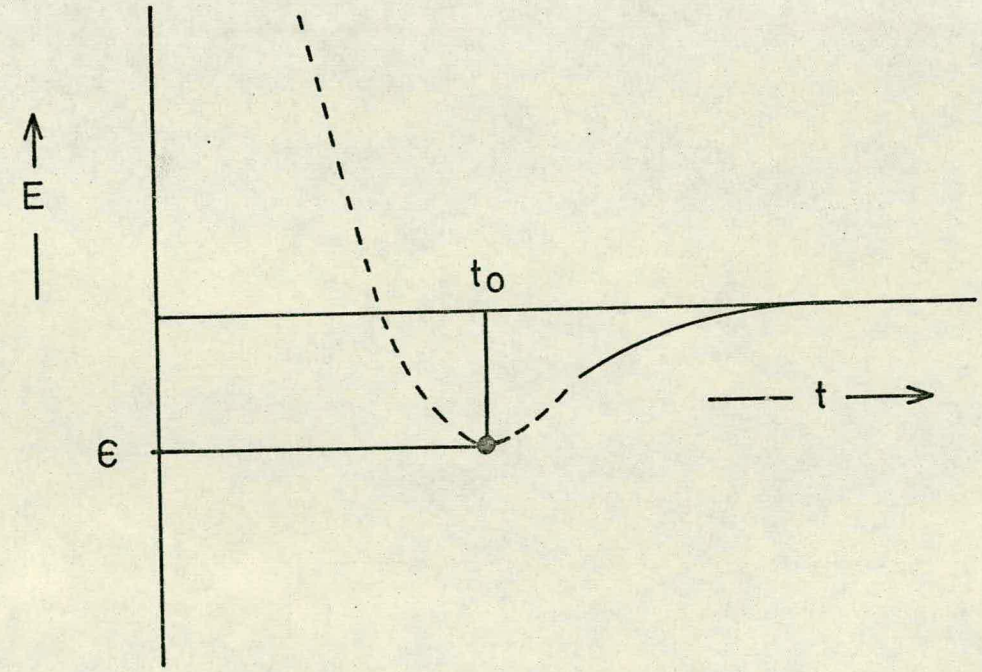


Fig. 19. The basis of the postulated function.

Consider the situation where $\theta = 0$

(θ, r, s, t as in fig. 18). Suppose the total interaction energy to be

$$V = at^{-9} - bt^{-6} + ct^{-1} + S(r) \qquad (1)$$

where a and b are to be determined, c is the usual polar attraction between O' and H and S(r) comprises all the other energy terms in the molecule. When $r = r_o$, take $t = t_o = r_o - s$, then

$$\epsilon = at_o^{-9} - bt_o^{-6} + ct_o^{-1} + S(r_o) \qquad (2)$$

and because this is a minimum, $\left(\frac{\partial V}{\partial t}\right)_\epsilon = 0$:

$$0 = -9a\, t_o^{-10} + 6b\, t_o^{-7} - c\, t_o^{-2} + \left[\frac{\partial}{\partial t} S(r)\right]_{r_o} \qquad (3)$$

This last term is difficult to evaluate, but it may be approximated as the derivative of the OO' interaction, T(r).

This will be

$$T(r) = dr^{-9} - er^{-6} + fr^{-1} \qquad (4)$$

with d, e and f as in the normal potential function, thus

$$\frac{\partial S}{\partial t} \doteq \frac{\partial T}{\partial t} = \frac{\partial T}{\partial r} = -9dr^{-10} + 6er^{-7} - fr^{-2} \qquad (5)$$

Put $\alpha = S(r_o)$ and $\beta = \left(\frac{\partial T}{\partial r}\right)_{r_o} \doteq \left(\frac{\partial S}{\partial t}\right)_{t_o}$; these are both known. Thus,

from (2), $$at_o^{-9} - b\, t_o^{-6} = \epsilon - \alpha - c\, t_o^{-1} \qquad (6)$$

from (3), $$-9a\, t_o^{-10} + 6b\, t_o^{-7} = -\beta + c\, t_o^{-2} \qquad (7)$$

whence $a = \frac{1}{5} (5\, ct_o^{8} + 6\,(\alpha - \epsilon)\, t_o^{9} + \beta t_o^{10})$ (8)

and $b = \frac{1}{5} (8\, ct_o^{5} + 9\,(\alpha - \epsilon)\, t_o^{6} + \beta t_o^{7}$ (9)

Once ϵ and t_o have been decided, a and b are thus determined.

For $\theta \neq 0$, the same function is used, but because $r < s + t$, the term S in (1) increases and the minimum is no longer as deep as ϵ. Fig. 20 shows the form of the potential. In this, as in the calculations to be described, ϵ is taken as -4.0 kcal/mole and r_o as 2.7 Å. It can be seen that as θ increases the minimum becomes shallower and moves to higher t, and vanishes beyond about $\theta = 60^{o}$.

In practice, this function is evaluated relatively rapidly. In a system with several possible bonds, a and b are different for each because of the polar terms in S, and must either all be evaluated first; or else more simply the coefficients of the polar terms in a and b, and the other terms, can be calculated and a and b generated in only a few operations each time.

Care must be taken that no interactions are counted twice: i.e. the O'O and O'H interactions must be excluded from the normal summations.

2.6 AVERAGING VERSUS MINIMISATION

If a system can exist in several freely interconvertible states of different energy then, in general, a

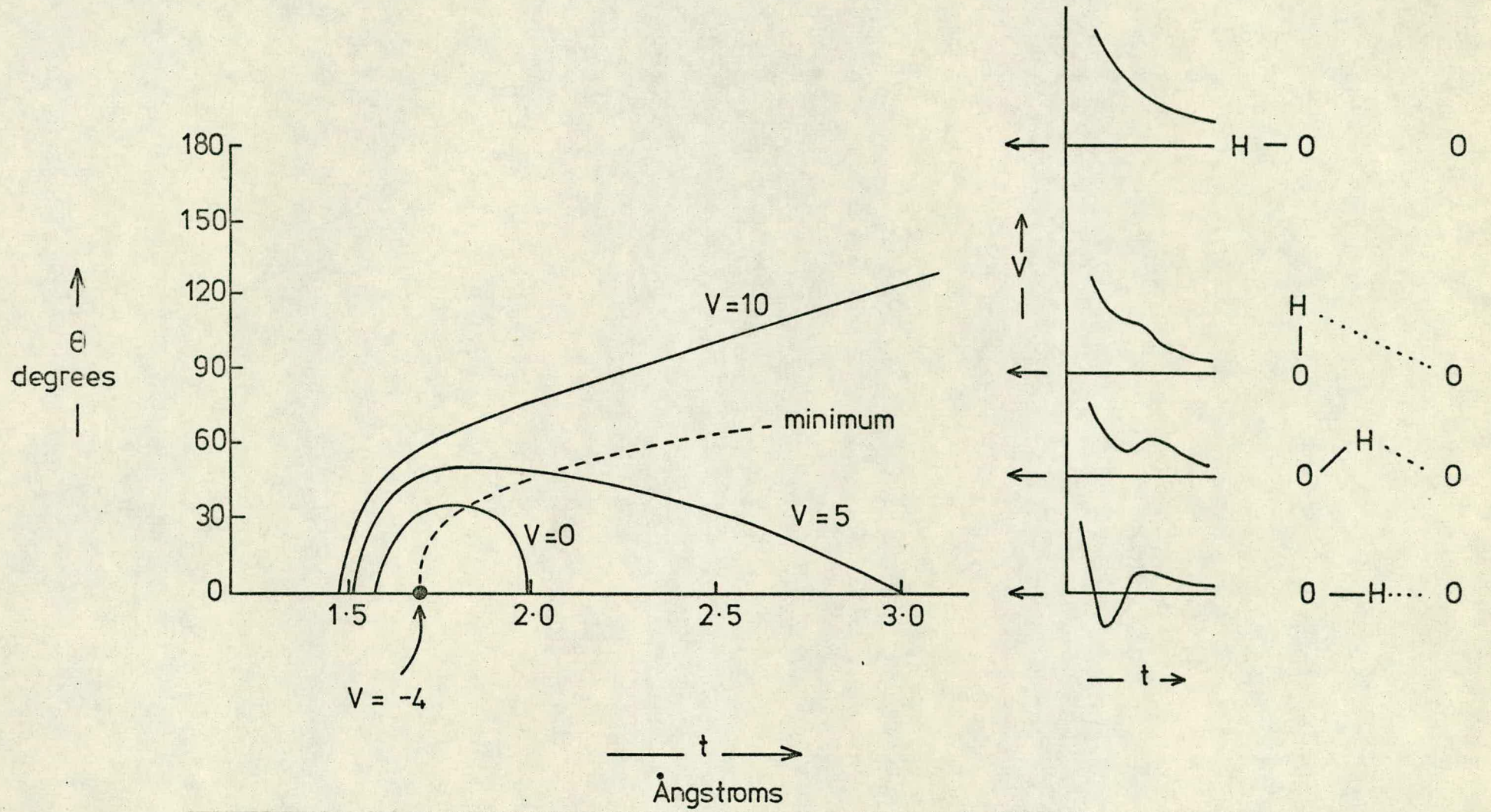


Fig. 20. H-bond potential with t_o = 1.7 Å, ϵ = - 4.0 kcal/mole.

'canonical ensemble', [117] of many such systems will exist such that the energy of the ensemble is partitioned between the systems according to the Maxwell-Boltzmann distribution. This predicts that the relative probability of any system existing in a state of energy ϵ_i when the ensemble is at equilibrium is $\exp(-\beta\epsilon_i)$ where $\beta=1/kT$, k being the Boltzmann constant and T the absolute temperature of the ensemble; or the absolute probability P_i is given by

$$P_i = \exp(-\beta\epsilon_i) \Big/ \sum_{\text{all states}} \exp(-\beta\epsilon_j). \qquad (1)$$

The average energy of any system is thus

$$\hat{\epsilon} = \frac{1}{N} \sum \epsilon_i \exp(-\beta\epsilon_i) \Big/ \sum \exp(-\beta\epsilon_i) \qquad (2)$$

where N is the number of states.

The conformational problem, in solution at least, concerns just such an ensemble: an assembly of molecules, each of which can exist in several easily interconvertible (thus ensuring equilibrium) states of different energy, and it follows that eqn. (2) applies. In principle, of course, there exists an infinity of states generated by all combinations of the infinitely variable conformation angles; in practice, however, a good approximation may be made in the usual way by varying the angles in discrete steps of sufficiently small size.

In the solid state, on the other hand, conformational

states are not normally readily interconverted, and further, in crystalline or other ordered systems, packing and lattice effects may combine with the intramolecular forces to favour certain conformations to the exclusion of others. The same may be true even in certain solutions, where ordered conformations (such as helices) persist. In these situations, the conformational energy is that of the particular conformation involved, or within the constraints imposed, the minimum energy conformation.

Even for 'free' solution work, however, few workers have adopted an averaging approach. Most have assumed that the energy of the minimum energy conformation is still the appropriate value, and have thus consistently underestimated the true average energy. Although this is clearly a source of considerable error, a certain amount of error cancellation will result from the subtraction involved in finding the energy difference between two conformations, the quantity normally of interest, and this particularly so if the energy wells are of the same shape (which, in general, they will not be). Minimisation is, moreover, easier and faster than averaging. Many procedures exist [118] for optimising a function of many variables, several of which have been reviewed with emphasis on their usefulness for conformational work by Gibson and Scheraga.[105] Appendix 1 contains a brief survey together with details of one method used extensively for various purposes in this work.

For the conformational energy problem, however, an attempt has been made to calculate averaged energies. If x_i, i = 1 to n are the conformational variables for a system, and $V(x_1, \ldots x_n)$ is the calculated potential function, then the problem reduces to the calculation of V for every combination of x values. Thus, for example, in a pentose with four conformation angles stepped, say, at 10^o intervals, there are $36^4 = 1.7 \times 10^6$ states whose energy must be calculated. Unfortunately, this is not normally practicable: about 10^4 evaluations is the practical limit if inordinate amounts of computer time are not to be used.

For this reason, a sampling technique must be used, whereby sets of values of $x_1 \ldots x_n$ are selected by some criterion and the average energy thereby approximated. The sampling is usually done randomly and the minimum size of the sample must be determined by experiment. In practice, this Monte Carlo method [119,120] reduces the number of necessary evaluations at the expense of accuracy. This method, however, suffers from the disadvantage that many of the sampled conformations may be of high energy and contribute little or nothing to the average. The ratio of the contributions from states of energies ϵ_i and ϵ_j is $\exp(-\beta(\epsilon_i - \epsilon_j))$. If ϵ_j is the lowest energy state, table 4 shows how the contribution from ϵ_i falls off as it increases, at 25°C to about 1% at 2.7 kcal/mole and 10^{-6} at 8.1 kcal/mole.

Table 4

Probability factors

ϵ kcal/mole	$\exp(-\epsilon/kT)$ at 0°C	at 25°C
0	1	1
0.1	0.83	0.84
0.2	0.69	0.71
0.5	0.40	0.43
1.0	0.16	0.18
2.0	0.025	0.034
5.0	9.6×10^{-5}	2.1×10^{-4}
10.0	9.2×10^{-9}	4.4×10^{-8}

Metropolis et al [121] have described an alternative sampling technique which increases the proportion of lower energy states selected. In this method, a state is chosen with a probability governed by its energy, and the energies of the states thus chosen are simply averaged: if the energies are ϵ_i', then $\hat{\epsilon} = \frac{1}{N} \sum \epsilon_i'$.

The procedure involves a first order Markov chain[120]: the probability of selection of a state, P_i, is not a priori constant, but is dependent upon the preceding selected state, so that there is a 'transition probability', P_{ij}, of going from state i to state j, which is independent of any states preceding i. If x_{ip}, i = 1 to n are the conformation variables after step p and δx_i, i = 1 to n are suitable (see later) steplengths for each variable, then step p+1 is generated as follows:

Let s, r_1, r_2 ... be random numbers, $-1 \leqslant r \leqslant 1$ and $0 \leqslant s \leqslant 1$. (see Appendix 2).

(1) Form trial $x_{i\ p+1} = x_i + r_i \delta x_i$

(2) Calculate ϵ_{p+1}, and $\Delta\epsilon = \epsilon_{p+1} - \epsilon_p$.

(3) If $\Delta\epsilon < 0$, new conformation is accepted.

(4) If $\Delta\epsilon > 0$, calculate $p = \exp(-\beta \Delta\epsilon)$.

(5) If $p > s$, new conformation is accepted.

(6) If $p < s$, new conformation is rejected, and the

$x_{i\ p+1}$ are given the values of x_{ip}, and ϵ_{p+1} that of ϵ_p.

Metropolis et al [121] have shown that this is in fact equivalent to choosing conformations with the appropriate probability.

Care must be taken in two places. Firstly, the starting conformation must be in a low energy region, which can be achieved by disregarding the first few steps until the system has settled down. Secondly, the δx_i must be chosen so that they are not too big so that too many conformations are rejected, and not too small, so that the system changes too slowly. A useful rule-of-thumb criterion is that about 20-30% of steps should be rejected.

The Metropolis method, like the plain Monte-Carlo method previously described, generates conformations which are in a specified manner representative of the ensemble as a whole. The Metropolis method may however be entirely equivalently thought of as a dynamic model (with unspecified timescale): i.e. as describing the behaviour of one molecule over a period of time, in which description the averaging is a simple time-average of the energy.

Averaging, then, is more complicated than minimisation in practice, and because it involves more evaluations of the energy function than many minimisation methods, uses more computing time. On the other hand,

it is undoubtedly a fundamentally more exact method for solution work and avoids the need to consider subsidiary energy minima which plague most minimisation methods (see Appendix 1).

2.7 CALCULATION PROCEDURE

Sections 2.2 to 2.5 have detailed the formulation of a potential function which may be applied to the calculation of carbohydrate conformational energies by the method of section 2.6.

The application of this to monosaccharides, disaccharides and, to a limited extent, polysaccharides will be described in chapters 3, 4 and 5 respectively. In the first two of these, the method is essentially similar and will be described in outline here. For disaccharides, the energy is calculated for each linkage conformation, so that the calculation to be described is carried out once for a monosaccharide and once for each investigated linkage conformation for a disaccharide.

In this energy calculation, atoms are classified as 'fixed' or 'floating', the former being those whose positions are not affected by the conformation angles and the latter those that are. There are thus three types of interaction, 'fixed-fixed' which are constant and thus left out of the Monte-Carlo cycle, 'fixed-float' and 'float-float' which are not. Fig. 21 is a flowchart of the calculation.

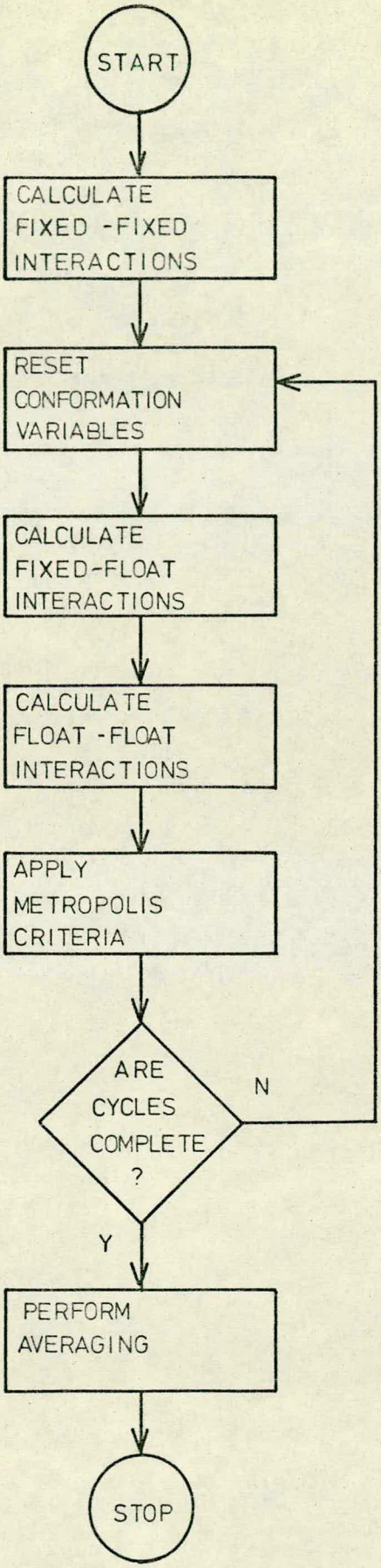


Fig. 21. Flowchart of energy calculation.

2.8 ENVIRONMENT EFFECTS

The method as outlined so far is strictly applicable only to finding the average energy of one isolated molecule *in vacuo*. In fact, the molecule will always of course be surrounded by either dissimilar molecules (in solution) or other similar ones (in the solid state). Unfortunately, it is very difficult to allow for this in a general way in calculations. A few specific allowances can however be made.

Intramolecular hydrogen bonding is clearly energetically less likely in a solvent such as water, where the water molecules, being much more mobile, will almost certainly be able to form better (in the sense of section 2.5) H-bonds and do so for a greater range of conformation angles, but this may be offset by the entropy lost in locally structuring the water molecules. Although the absence of such intramolecular bonds is easily allowed for, the entropy term and the possibility that certain conformations may crowd out solvent molecules from forming H-bonds are not.

The dielectric constant, especially at large interatomic separations, is obviously affected by the solvent. Since, however, its value is established largely empirically, it is not obvious how it should be altered (see section 2.3 and Ramachandran's paper [108]).

Gibson and Scheraga [105] have in peptide work derived a method for estimating the free energy of solvation which expresses quantitatively the tendency for polar groups to lie on the surface of a polymer in a polar solvent, and non-polar groups to lie in the interior of the molecule.

For interactions with similar molecules, x-ray diffraction work (see section 1.5) combined with calculation methods often enable the relative positions of polysaccharide strands in helical polymers to be established. Usually, however, a full potential function treatment is impracticable and hybrid methods must be employed (see chapter 5).

Undoubtedly, environment effects are important, and the lack of any available method for their full treatment is one contributory factor to the remaining uncertainty of the structures of, for example, cellulose and amylose.

2.9 ENTHALPIES AND FREE ENERGIES

The energy functions mentioned so far are all concerned with potential energy, i.e. enthalpy, differences, ΔH, between systems. Measured energy differences, however, are usually of free energy, ΔG. These are related by

$$\Delta G = \Delta H - T \Delta S \qquad (1)$$

where ΔS is the entropy difference and T the absolute temperature.

The assumption usually made is that ΔS is negligible: that is, that the number of conformational energy states available to the two species considered is the same. Brant et al [98] have calculated conformational entropies on the basis of summed partition functions similar to those discussed in section 2.6, but over linkage conformations in polypeptides. They point out, however, that their values include an undetermined additive constant because of the approximation of summation for integration. The differences between the values they obtained were small: i.e. $T\Delta S$ was much less than ΔH, and the $\Delta G \approx \Delta H$ approximation thereby justified for their system.

Vijayalakshmi and Rao [103] have recently published a primitive and tentative evaluation of conformational entropies in monosaccharides. This is based on energy calculations which will be discussed in chapter 3.

2.10 SUMMARY

A potential function has now been postulated containing two terms concerning two atom interactions, van der Waals and polar; and two concerning more complex interactions, torsion and hydrogen bonding. A method has been described for the application of this function, which will be detailed in the following chapters.

CHAPTER 3

THE CONFORMATIONAL ENERGY OF MONOSACCHARIDES

3.1 INTRODUCTION

A great deal of effort has gone into the study of monosaccharide conformations, and for this reason much qualitative and some quantitative evidence exists as to their relative energies.

The work described in this chapter seeks to calculate conformational energies for the pentopyranose sugars. These have been studied in preference to hexoses because of the complications introduced by the extra degree of freedom in the latter, due to rotation about the C(6)-O(6) bond.

With a C1 chair ring form, pentoses may have axial or equatorial hydroxyl groups at each of C(1), C(2), C(3) and C(4), thus giving sixteen possible isomers. Not all of these exist in this chair form, but for comparison purposes all sixteen have been thus taken.

For this purpose, each pentose may then be identified by indicating the configuration of the OH groups at C(1), C(2), C(3) and C(4). Table 5 shows for all sixteen how this relates to the conventional nomenclature, using E for equatorial and A for axial; thus EEEE is β-D-xylose, AEEE is α-D-xylose and so on. The symbol 'x' will be used to signify 'either', thus xEEE is α- or β-D-xylose. This nomenclature will be used throughout this chapter.

Table 5

Nomenclature of C1 chair pentoses and correspondence of 1C mirror images.

A = axial hydroxyl, E = equatorial hydroxyl

Configuration at carbon 1	2	3	4	Conventional names: actual (C 1)	mirror image (1C)
A	E	A	E	α-D-ribose	corresponding L sugars, e.g. α-L-ribose
E	E	A	E	β-D-ribose	
A	A	A	E	α-D-arabinose	
E	A	A	E	β-D-arabinose	
A	E	E	E	α-D-xylose	
E	E	E	E	β-D-xylose	
A	A	E	E	α-D-lyxose	
E	A	E	E	β-D-lyxose	
E	A	E	A	α-L-ribose	corresponding D sugars, e.g. α-D-ribose
A	A	E	A	β-L-ribose	
E	E	E	A	α-L-arabinose	
A	E	E	A	β-L-arabinose	
E	A	A	A	α-L-xylose	
A	A	A	A	β-L-xylose	
E	E	A	A	α-L-lyxose	
A	E	A	A	β-L-lyxose	

Although this nomenclature relates to the C1 chairs, the molecules thus denoted are mirror images of 1C chairs with, of course, the same conformational energies; thus 1C β-D-xylose is the mirror image of C1 β-L-xylose and so on. These equivalences are also shown in table 5.

For nine of these pentoses, no major interaction exists which might be expected to force them out of the C1 conformation. The remaining seven, which are xAxA and AxAx, have, however, a strong repulsion between the two axial oxygen atoms, and depending upon the configuration at the other carbon atoms, may occur naturally 'flipped' to a 1C chair or otherwise distorted. These latter will be denoted here 'strained' and the remaining nine 'unstrained'. The unstrained pentoses are: EEEE, AEEE, EAEE, EEAE, EEEA, AAEE, AEEA, EAAE and EEAA.

An energy function, such as has been postulated in the preceding chapter, containing no terms for angle strain or bond stretching is most directly applicable to the unstrained molecules: calculations on these are described in section 3.5, and on the strained ones in section 3.6.

Angyal [12,13,36] has systematised much of the experimental work in this field and has devised empirical rules for predicting the conformational free energies of

monosaccharides: his methods are discussed in section 3.2 and later used as a criterion of success for the calculations.

Rao and others [93,103,106,107] have recently published work comparable to that attempted here: this is considered in section 3.3.

3.2 THE METHOD OF ANGYAL

Angyal [12,13] assumes certain additive contributions to the conformational energy of pyranose sugars which arise from interactions as follows, insofar as they relate to pentoses:

(a) between axial atoms on next-to-adjacent carbon atoms, unless both are hydrogen;

(b) between atoms on adjacent carbon atoms if neither is hydrogen and at least one is equatorial; and

(c) the 'anomeric effect' for equatorial C(1), the value of which depends on the configuration and substituent at C(2).

He assigns values to these interactions as follows:

(a) O-H, 0.45; O-O, 1.50 kcal/mole;

(b) O-O, 0.35 kcal/mole; and

(c) for EExx, 0.55; for EAAx, 0.85; and for EAEx, 1.00 kcal/mole. The values for the sixteen pentoses are given in table 6.

Table 6

Angyal free energies for the pentoses

Pentose	Angyal energy kcal/mole (a)	(b)	(c)	total (i)	(ii)
Unstrained:					
E E E E	0	1.05	0.55	1.60	0
A E E E	0.90	1.05	0	1.95	0.35
E A E E	0.45	1.05	1.00	2.50	0.90
E E A E	0.90	1.05	0.55	2.50	0.90
E E E A	0.45	1.05	0.55	2.05	0.45
A A E E	1.35	0.70	0	2.05	0.45
A E E A	1.35	1.05	0	2.40	0.80
E A A E	1.35	0.70	0.85	2.90	1.30
E E A A	1.35	0.70	0.55	2.60	1.00
Strained:					
A E A E	2.40	1.05	0	3.45	1.85
E A E A	1.50	1.05	1.00	3.55	1.95
A A A E	2.85	0.35	0	3.20	1.60
A A E A	2.40	0.70	0	3.10	1.50
A E A A	2.85	0.70	0	3.55	1.95
E A A A	2.40	0.35	0.85	3.60	2.00
A A A A	3.90	0	0	3.90	2.30

Total (ii) is relative to EEEE

These values are free energies in aqueous solution and were mostly obtained from the measurement of equilibrium constants, which give direct information about free energy differences, since for two species in equilibrium, if the equilibrium constant is K,

$$\Delta G^{o} = -RT \ln K \qquad (1)$$

where ΔG^{o} is the standard free energy difference between the species, T the absolute temperature, and R the gas constant.

The methods to be described have been compared with these Angyal energies, and as a measure of agreement, the standard deviation from the Angyal values has been calculated. Thus, for n pentoses,

$$\sigma = \sqrt{\frac{1}{n} \sum (\epsilon_i - \alpha_i)^2} \qquad (2)$$

where ϵ_i and α_i are the calculated and Angyal values respectively for the i th pentose. This value is statistically uncertain due to the small size of n, and has, because of this, a standard error of $\sigma / \sqrt{2(n-1)}$, or about 35% for n = 9 and 18% for n = 16.

3.3 THE METHOD OF RAO

Rao and co-workers have performed calculations for hexoses and pentoses without polar terms[106], and for hexoses [93,107] and pentoses [103] with polar terms.

The potential function they used was in three parts, van der Waals, polar, and angle strain.

The van der Waals potential was that of Kitaigorodsky (see section 1.11):

$$V_{ij} = 3.5\ (8600\ \exp(-13Z_{ij}) - 0.04/Z_{ij}^{6}) \qquad (1)$$

where $Z = r_{ij}/r_o$. The values of r_o were usually those given by Rao et al[78], although for some of calculations[107] the radius of the oxygen was reduced by about 0.1 A, to improve the agreement of the results.

The polar potential was the same as used in the present work, but with the effective dielectric constant variously 3.5 for solution [93,103] and 7.0 for solution and 2.0 for solid state [107], the charges used were the same as already quoted, except for insignificant deviations in one investigation [107].

The angle strain potential was harmonic:

$$V = k(\theta - 109.5)^2 \qquad (2)$$

where θ is the bond angle in degrees, and the force constants, k, were estimated from spectroscopic data. The only angles allowed to vary were those involving pendant atoms, so that at each ring carbon atom, five angles were allowed to deform. Because of the assumptions used (see below) in calculating the coordinates, these angles were all initially tetrahedral.

The atom coordinates used were based on a set of 'ideal' bond lengths and angles of which further details are given in the following section.

In each calculation the energy was minimised by an iterative procedure [122]. This consisted of, from an arbitrary starting conformation, fixing each conformation angle in turn at its minimum energy position, continuing the process in a cyclic fashion until the process converged at what was assumed to be the global minimum energy conformation.

In fact, this assumption is incorrect. The process will infallibly find a global minimum only if each rotation gives a single minimum which is primarily determined by the core of the molecule and only slightly perturbed by the other rotating groups. In Rao's system, this is probably true for the hydroxyl groups on the ring carbons, but is not true for the rotations fixing O(6) and H(O6), as can be shown by repeating the hexose work but rotating either of these atoms in the opposite sense. Fortunately, although the conformational angles at minimum energy are very different, the energy is not.

The fallacy in the method can best be illustrated by a hypothetical pathogenic case. Suppose a system to have four conformation angles, θ_1 to θ_4, and that the energy of the system is -10 units for only $\theta_1=\theta_2=\theta_3=\theta_4 = 180^\circ$,

-1 unit when any $\theta = 90^{o}$, and zero otherwise. Then Rao's method will miss the global minimum for any starting values except those with at least three angles already 180^{o}.

For pentoses, where no atoms rotate about other rotated atoms, Rao's method is probably successful in locating the overall minimum, but only because he omits a torsional potential, which would lead to more pronounced subsidiary minima.

The reason given for omitting the torsional effect from the potential is that better agreement is obtained without. This, however, is very difficult to reconcile with other already considered work on the torsional effect, and unless a satisfactory explanation can be found, must be considered a serious flaw in the method.

Rao's results for pentoses are given in tables 7 and 8. It is worth noting that his methods were developed on hexoses, for which agreement with Angyal was somewhat better, and that this loss of predictive power in applying the technique to the very similar pentoses suggests a possible lack of generality. He also attempts better to match Angyal's free energies by including a crudely estimated entropy term (see sections 2.9 and 3.7): this is not quoted here.

Table 7

Energies in kcal/mole calculated by Rao

CONF	(1) NORM	(2) DIST	(3) ANG
EEEE	-.72	-.59	-.68
AEEE	-.12	-.09	-.33
EAEE	.28	.41	.22
EEAE	-.16	-.06	.22
EEEA	-.04	.09	-.23
AAEE	.32	.34	-.23
AEEA	.41	.54	.12
EAAE	.31	-.50	.62
EEAA	-.28	-.17	.32

UNSTRAINED PENTOSES - RAO'S RESULTS

(1) Without angle strain, σ = 0.35 kcal/mole
(2) With angle strain, σ = 0.50 kcal/mole
(3) Angyal values.
σ is the standard deviation from the Angyal values

Table 8

Energies in kcal/mole calculated by Rao.

CONF	(1) NORM	(2) DIST	(3) ANG
EEEE	-1.42	-1.11	-1.21
AEEE	-.82	-.61	-.86
EAEE	-.42	-.11	-.31
EEAE	-.86	-.58	-.31
AEEE	-.74	-.43	-.76
AAEE	-.38	-.18	-.76
AEEA	-.29	.02	-.41
EAAE	-.39	-1.02	.09
EEAA	-.98	-.69	-.21
AEAE	.85	.80	.64
EAEA	1.32	1.19	.74
AAAE	.50	.45	.39
AAEA	.84	.70	.29
AEAA	.62	.57	.74
EAAA	.68	.30	.79
AAAA	1.49	.75	1.09

ALL PENTOSES - RAO'S RESULTS

(1) Without angle strain, σ = 0.37 kcal/mole.
(2) With angle strain, σ = 0.44 kcal/mole
(3) Angyal values.
σ is the standard deviation from the Angyal values.

The results are given for the unstrained, and for all pentoses. They are adjusted in each set so that the mean of the set is zero, as are the Angyal values, so that the standard deviation is a minimum.

3.4 ATOM COORDINATES

Most published determinations of molecular dimensions are in terms of bond lengths and angles. When the molecule contains a closed ring, the calculation thence of cartesian coordinates of the atoms is not mathematically trivial.

For a six-membered ring, there exist three degrees of freedom per atom giving eighteen in all, less three arbitrarily fixing it in space and three orienting it with respect to the coordinate frame. There are thus twelve independent parameters, and these are to be obtained from six bond lengths and six bond angles: the system is thus fully determined and will have a limited number of exact solutions. There may be none (if, e.g., all the bond angles are 180°) or, in a real situation, two different chairs and possibly some boats.

The following method was developed for the computation of these ring atom coordinates, by refining an initial 'trial' ring which approximates to the desired one.

(i) Let P_1 to P_6 define a suitable trial ring (see below) with coordinates (x_1, y_1, z_1) etc., bond lengths

s_1 (= P_1P_2) to s_6, bond angles

b_1 (= $P_6P_1P_2$) to b_6.

(ii) Let r_1 to r_6 and a_1 to a_6 be respectively the desired lengths and angles.

(iii) Minimise the function

$$F = \sum (r_i - s_i)^2 + \sum (a_i - b_i)^2 \qquad (1)$$

with respect to the 18 variables, x_1, y_1, z_1, x_2 etc. For an exact solution, the minimum F will be zero, furthermore, if the trial ring is chosen to be in the desired conformation (e.g. C1 chair) and the step size used in the minimisation procedure is sufficiently small, then the final ring will be an arbitrarily placed and oriented ring with the desired properties.

A computer program was written to do this using the Nelder and Mead simplex method (see Appendix 1). The function F was modified so that the terms were all about the same magnitude when the lengths were in Ångstroms and the angles in degrees, by dividing the angle terms by 70. Steplengths were taken as 0.1 (Å or x70°) and the procedure converged to $F < 10^{-8}$ after about 2000 function evaluations. Agreement between 'asked' and 'found' parameters was to about 1 in 10^5.

Two further steps in this program added pendant groups and oriented the whole.

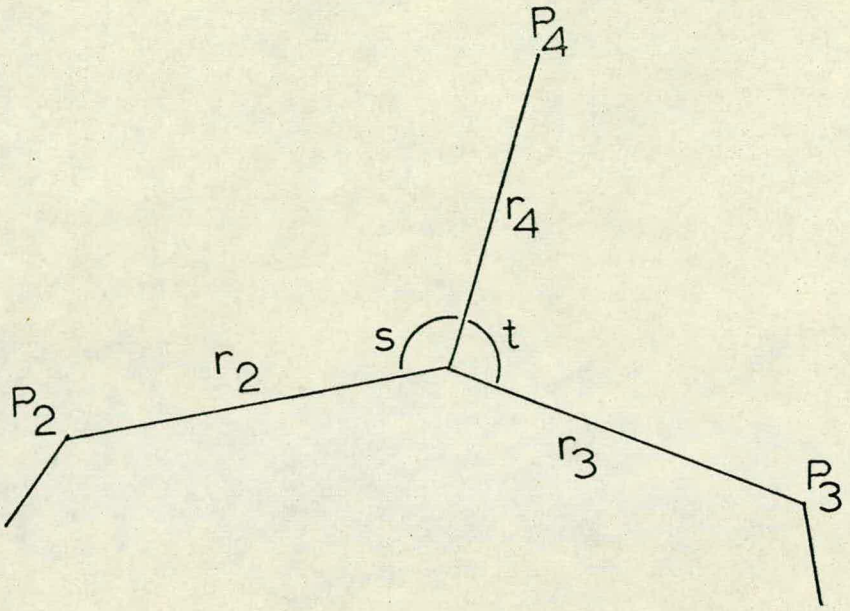


Fig. 22. Pendant atom parameters

The position of a pendant atom is fully defined (see fig. 22) by:

(a) the bond length from it to the ring atom, r_4,

(b) the angles between this bond and the ring bonds backwards and forwards, s and t respectively, and

(c) whether it is axial or equatorial. The two positions meeting criteria (a) and (b) may be determined as follows: Let $P_2\ P_1\ P_3$ be the relevant part of the ring and P_4 be the new point to be established. Let x_j, y_j, z_j be the components of P_1P_j (see fig. 1);

thus $x_2x_4 + y_2y_4 + z_2z_4 = r_2r_4\cos s$ (2)

$x_3x_4 + y_3y_4 + z_3z_4 = r_3r_4\cos t$ (3)

and $x_4^2 + y_4^2 + z_4^2 = r_4^2$ (4)

Let the right hand sides of (2) and (3) be A and B respectively.

Thus from (2) $x_2x_3x_4 + x_3y_2y_4 = x_3 (A-z_2z_4)$ (5)

and from (3) $x_2x_3x_4 + x_2y_3y_4 = x_2 (B-z_3z_4)$ (6)

whence $y_4 \quad \dfrac{x_3(A-z_2z_4) - x_2 (B-z_3z_4)}{x_3y_2 - x_2y_3}$ (7)

Let the denominator be C.

$\therefore \quad y_4 = \left[(x_3A - x_2B) / C \right] + \left[(-x_3z_2+x_2z_3) /C \right] Z_4$ (8)

$= D + E Z_4$, say (9)

Similarly, $x_4 = F + GZ_4$ (10)

Where $F = (y_3A - y_2B) / H$ (11)

$G = (-y_3z_2 + y_2z_3) / H$ (12)

and $H = y_3x_2 - y_2x_3$ (13)

Thus from (4), (9) and (10)

$(F + G z_4)^2 + (D+E z_4)^2 + z_4^2 = r^2$ (14)

and if the polynomial coefficients of z_4 are

$a = G^2 + E^2 + 1, \quad b = 2FG + 2DE, \quad c = F^2 + D^2 - r_4^2$ (15)

the two values of z_4 are given as usual by

$$z_4 = \left[-b \pm \sqrt{(b^2 - 4ac)} \right] / 2a \qquad (16)$$

The square root will always be real for a valid problem.

The two values of x_4 and y_4 can thus be obtained from (9) and (10) and the appropriate point chosen by consideration of the fact that the axial position is closer to a ring atom next but one to its attached one (e.g. axial O(1) is nearer C(3) than is equatorial O(1)).

The only singularity that can arise is if C (or H, which is -C) is zero or close to zero. Now $C = x_2y_3 - x_3y_2 = -q_2q_3 \sin x$ where q_i is the projection of r_i on the xy plane, and x the angle between q_2 and q_3. This will be small if q_2 or q_3 is small, or x is close to 0° or 180°, that is, when the plane of $P_1P_2P_3$ is nearly perpendicular to the xy plane.

The procedure adopted to overcome this was to test if $|x_2| + |x_3|$ was less than 0.2, and if so to permute all the coordinates thus: $z \leftarrow x$, $x \leftarrow y$, $y \leftarrow z$, or if $|y_2| + |y_3|$ was less than 0.2 to permute: $z \leftarrow y$, $y \leftarrow x$, $x \leftarrow z$. This is simply done, and involves no loss of precision due to rounding errors such as other methods might, and it is easily seen that this modification will cope with even the worst possible orientation of the molecule.

The last step is to orient the molecule with respect to the axes. This requires the user to specify

(i) an atom to be the origin,

(ii) an atom to lie on a specified semi-axis, and

(iii) an atom to lie in a quarter-plane formed by this and another specified semi-axis.

This entails one translation to the origin and three rotations. Details of this frequently required operation are given in Appendix 3.

General programming details are given in Appendix 4.

Arnott and Scott[15] have subsequently adopted a very similar approach which differs only in including the ring dihedral angles (e.g. that between the planes C(1)-C(2)-C(3) and C(2)-C(3)-C(4)) in the minimisation. This gives an over-determined system in which F will not necessarily minimise to zero, and the weighting factors assigned to lengths, bond angles and dihedral angles will be important in determining the result. In fact Arnott and Scott give low significance to the dihedral angles so that the values obtained differ only slightly.

For the present work, two sets of coordinates have been used, denoted 'ideal' and 'A & S'. The ideal set were calculated as above from mean values taken primarily from Settineri and Marchessault's work [23]; the A & S set are from Arnott and Scott's paper, and are mean values calculated by their method from a large number of crystallographic determinations.

Table 9

Comparison of bond length and angle assumptions

			Ideal	A & S	Rao
Ring:	C(1)-C(2)		1.54	1.523	1.53
	C(2)-C(3)		1.54	1.521	1.53
	C(3)-C(4)		1.54	1.523	1.53
	C(4)-C(5)		1.54	1.525	1.53
	C(5)-O(5)		1.42	1.436	1.53
	O(5)-C(1)	(a)	1.42	1.429	1.42
		(b)	1.42	1.414	1.42
	O(5)-C(1)-C(2)		110.0	109.2	109.47
	C(1)-C(2)-C(3)		110.0	110.5	109.47
	C(2)-C(3)-C(4)		110.0	110.5	109.47
	C(3)-C(4)-C(5)		110.0	110.3	109.47
	C(4)-C(5)-O(5)		110.0	110.0	109.47
	C(5)-O(5)-C(1)	(c)	111.0	112.0	114.0
		(d)	111.0	114.0	114.0
Pendant:	C(1)-O(1)	(a)	1.43	1.389	1.42
		(b)	1.43	1.415	1.42
	C(2)-O(2)		1.43	1.426	1.42
	C(3)-O(3)		1.43	1.426	1.42
	C(4)-O(4)		1.43	1.426	1.42
	C(ring)-H		1.09	1.1	1.10
	O(5)-C(1)-O(1)	(c)	110.0	107.3	109.47
		(d)	110.0	111.6	109.47
	C(2)-C(1)-O(1)		110.0	108.4	109.47
	C(1)-C(2)-O(2)		110.0	109.3	109.47
	C(3)-C(2)-O(2)		110.0	110.8	109.47
	C(2)-C(3)-O(3)		110.0	109.6	109.47
	C(4)-C(3)-O(3)		110.0	109.7	109.47
	C(3)-C(4)-O(4)		110.0	110.4	109.47
	C(5)-C(4)-O(4)		110.0	108.6	109.47
	C/O(ring)-C(ring)-H		109.5	109.5	109.47

Notes:
(a) equatorial O(1) and axial O(1)-H
(b) axial O(1)-R with R ≠ H
(c) equatorial O(1)
(d) axial O(1)

Table 10

(a) Atomic coordinates for pentose - Ideal
The six groups are (i) ring, (ii) and (iii) pendant atoms for equatorial oxygen, (iv) and (v) pendant atoms for axial oxygen, (vi) H(5a) (equatorial) and H(5b) (axial).

	X	Y	Z
C(1)	.0000	.0000	.0000
C(2)	1.5400	.0000	.0000
C(3)	2.0668	1.4470	.0000
C(4)	1.3908	2.2536	-1.1243
C(5)	-.1383	2.0910	-1.0425
O(5)	-.4856	.7167	-1.1256
O(1)	-.4891	-1.3435	-.0274
O(2)	2.0290	-.6986	1.1479
O(3)	3.4854	1.4509	-.1802
O(4)	1.7423	3.6354	-1.0143
H(1)	-.3638	.4686	.9144
H(2)	1.9037	-.5197	-.8864
H(3)	1.8516	1.9126	.9618
H(4)	1.7435	1.8957	-2.0915
O(1)	-.4891	.5932	1.2057
O(2)	2.0290	-.6986	-1.1479
O(3)	1.8093	2.0613	1.2654
O(4)	1.8627	1.8072	-2.3982
H(1)	-.3638	-1.0268	-.0378
H(2)	1.9038	-.5197	.8863
H(3)	3.1458	1.4413	-.1544
H(4)	1.6506	3.3073	-1.0229
H(5A)	-.6064	2.6349	-1.8630
H(5B)	-.4978	2.5007	-.0986

(continued over)

Table 10 (continued)

(b) Atomic coordinates for pentose - A & S with axial O(1). Groups as in (a).

	X	Y	Z
C(1)	.0000	.0000	.0000
C(2)	1.5230	.0000	.0000
C(3)	2.0551	1.4249	.0000
C(4)	1.4465	2.2197	-1.1477
C(5)	-.0743	2.1117	-1.1144
O(5)	-.4652	.7300	-1.1181
O(2)	1.9943	-.7166	1.1392
O(3)	3.4752	1.4052	-.1288
O(4)	1.8071	3.5956	-1.0457
H(2)	1.8902	-.5291	-.8918
H(3)	1.8059	1.9099	.9553
H(4)	1.8215	1.8283	-2.1049
O(1)	-.4384	.5586	1.1938
O(2)	1.9943	-.7166	-1.1392
O(3)	1.7395	2.0534	1.2405
O(4)	1.9107	1.7195	-2.3998
H(1)	-.3672	-1.0352	-.0587
H(2)	1.8902	-.5291	.8918
H(3)	3.1500	1.4079	-.1051
H(4)	1.7416	3.2757	-1.0601
H(5A)	-.4976	2.6154	-1.9960
H(5B)	-.4572	2.5993	-.2057

(continued over)

Table 10 (continued)

(c) Atomic coordinates for pentose - A & S with equatorial O(1). Groups as in (a)

	X	Y	Z
C(1)	.0000	.0000	.0000
C(2)	1.5230	.0000	.0000
C(3)	2.0565	1.4244	.0000
C(4)	1.4407	2.2192	-1.1440
C(5)	-.0795	2.1052	-1.1051
O(5)	-.4733	.7246	-1.1371
O(1)	-.4384	-1.3112	-.1240
O(2)	1.9944	-.7172	1.1388
O(3)	3.4756	1.4037	-.1382
O(4)	1.7964	3.5962	-1.0405
H(1)	-.3672	.4791	.9196
H(2)	1.8902	-.5296	-.8915
H(3)	1.8113	1.9082	.9570
H(4)	1.8136	1.8309	-2.1033
O(2)	1.9944	-.7172	-1.1388
O(3)	1.7461	2.0514	1.2426
O(4)	1.9021	1.7230	-2.3988
H(2)	1.8902	-.5296	.8915
H(3)	3.1506	1.4067	-.1123
H(4)	1.7321	3.2762	-1.0553
H(5A)	-.5093	2.6298	-1.9712
H(5B)	-.4595	2.5742	-.1855

A comparison of ideal, A & S and Rao (see section 3.3) assumptions is given in table 9, and the ideal and A & S coordinates in table 10.

3.5 UNSTRAINED PENTOSES - RESULTS

The calculations as described in previous sections were first carried out on the unstrained pentoses. Several runs were performed to establish the best values for (a) the number of Monte-Carlo cycles, and

(b) the maximum step size for the Metropolis averaging (see section 2.6).

The latter was chosen by inspection of the progress of the Markov chain. After the initial settling down (which was excluded from the averaging), a step size was required which would cause a 20 to 30% rejection rate. Steps of 5°, 10° and 15° were tried, and the system was found not to react critically to this parameter: a value of 10° was chosen.

Runs of 500,1000, 2000, 5000 and 10000 cycles were then carried out using different random seeds (see Appendix 2). A value of 5000 was chosen as being the best compromise between convergence and economy, giving energies with a standard deviation from the mean of about 0.1 kcal/mole.

The torsional energy term being the least well characterised, and also (see section 3.3) having been omitted by other workers, a series of runs was performed

Table 11

Calculated energies in kcal/mole.
Best fit for A & S coordinates: 0.7 torsion.

CONF	(1) FIX	(2) FLOAT	(3) TOTAL	(4) Ø MN	(5) ANG	(6) RAO
EEEE	19.68	-13.64	6.04	.06	-.68	-.72
AEEE	20.27	-14.67	5.60	-.38	-.33	-.12
EAEE	21.75	-15.70	6.05	.07	.22	.28
EEAE	20.53	-14.53	6.00	.02	.22	-.16
EEEA	22.08	-16.13	5.95	-.03	-.23	-.04
AAEE	19.29	-13.98	5.31	-.67	-.23	.32
AEEA	21.37	-15.36	6.01	.03	.12	.41
EAAE	19.62	-12.87	6.75	.77	.62	.31
EEAA	19.78	-13.69	6.09	.11	.32	-.28

UNSTRAINED PENTOSES - A+S COORDINATES

(1) Fixed-fixed interactions.
(2) Fixed-float and float-float interactions
(3) Total
(4) Same, adjusted to have zero mean
(5) Angyal values
(6) Rao undistorted values (see section 3.3)

Table 12

Calculated energies in kcal/mole.
Best fit for ideal coordinates : 0.8 torsion

CONF	(1) FIX	(2) FLOAT	(3) TOTAL	(4) Ø MN	(5) ANG	(6) RAO
EEEE	19.49	-13.60	5.89	-.18	-.68	-.72
AEEE	19.96	-14.28	5.68	-.39	-.33	-.12
EAEE	21.28	-14.95	6.33	.26	.22	.28
EEAE	20.41	-14.10	6.31	.24	.22	-.16
EEEA	21.79	-15.69	6.10	.03	-.23	-.04
AAEE	23.37	-17.26	6.11	.04	-.23	.32
AEEA	21.08	-15.17	5.91	-.16	.12	.41
EAAE	19.09	-12.62	6.47	.40	.62	.31
EEAA	19.59	-13.79	5.80	-.27	.32	-.28

UNSTRAINED PENTOSES - IDEAL COORDINATES

Columns as in table 11.

Table 13

Standard deviation from Angyal values - unstrained pentoses.

Set	σ (kcal/mole)
A & S (table 1)	0.32
Ideal (table 2)	0.31
Rao - no angle strain	0.35
with angle strain	0.50
(Section 3.3)	

using different proportions of the assumed barrier height; these were 0, 0.3, 0.5, 0.7, 0.8, 0.9 and 1.

Both sets of coordinates were used for each. Fig. 23 shows a plot of the standard deviation from Angyal, σ, (see section 3.2) as a function of the torsional barrier, and it can be seen that a value of about 0.7 - 0.8 is best. The best fits for A & S and ideal coordinates are given in tables 11 and 12 respectively, showing the fixed and floating contributions (see section 2.7) and the comparison with Angyal and Rao values. Standard deviations from Angyal values are given in table 13 from which the method is shown to be marginally better than that of Rao.

3.6 STRAINED PENTOSES - RESULTS

The strained pentoses are those which contain two axial hydroxyl groups on the same side of the ring: they are AEAE, EAEA, AAAE, AAEA, AEAA, EAAA and AAAA. All but the first two of these can relieve this strain by flipping to the 1C chair conformation, and it is found experimentally that they do, with the exception of AAEA, which exists like AEAE and EAEA, as a mixture of both forms[123].

The calculation described previously was carried out on these pentoses for each set of coordinates using the best torsion barrier as established for the unstrained case. All other parameters were kept the same as before.

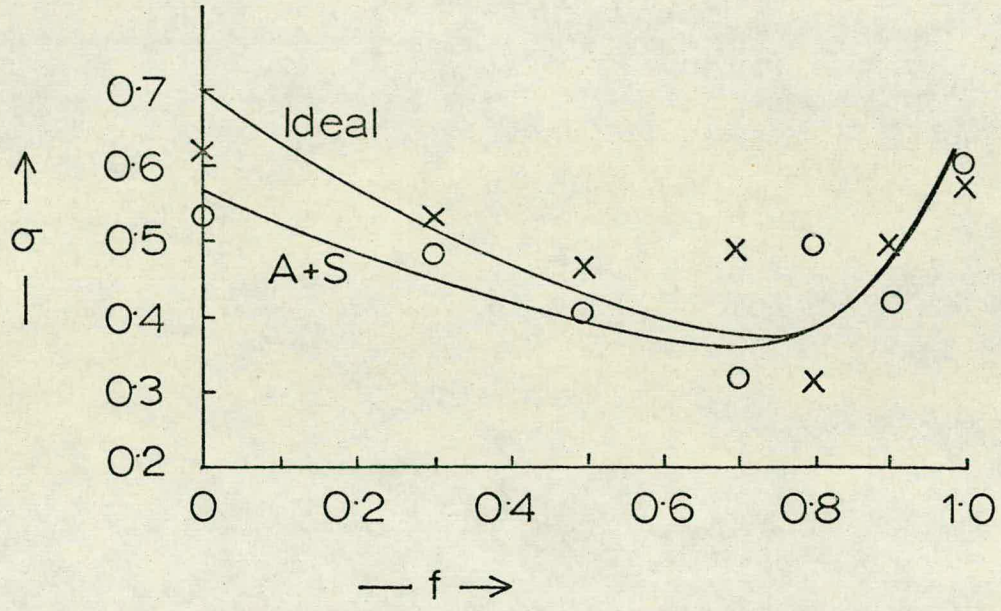


Fig. 23. Dependence of standard deviation from Angyal (σ, in kcal/mole) on value assumed for torsion barrier, fT, where T is the 'normal' value

x - ideal coordinates
o - A & S coordinates

The standard error in σ is about 35%.

Table 14

Calculated energies in kcal/mole.
Best fit for A & S coordinates : 0.7 torsion.

CONF	(1) FIX	(2) FLOAT	(3) TOTAL	(4) Ø MN	(5) ANG	(6) RAO
EEEE	19.68	-13.64	6.04	-.14	-1.21	-1.11
AEEE	20.27	-14.67	5.60	-.57	-.86	-.61
EAEE	21.75	-15.70	6.05	-.12	-.31	-.11
EEAE	20.53	-14.53	6.00	-.18	-.31	-.58
EEEA	22.08	-16.13	5.95	-.23	-.76	-.43
AAEE	19.29	-13.98	5.31	-.86	-.76	-.18
AEEA	21.37	-15.36	6.01	-.17	-.41	.02
EAAE	19.62	-12.87	6.75	.57	.09	-1.02
EEAA	19.78	-13.69	6.09	-.09	-.21	-.69
AEAE	24.36	-17.93	6.43	.26	.64	.80
EAEA	27.17	-19.85	7.32	1.14	.74	1.19
AAAE	20.53	-14.03	6.50	.32	.39	.45
AAEA	23.48	-17.53	5.95	-.22	.29	.70
AEAA	22.43	-16.41	6.02	-.16	.74	.57
EAAA	21.89	-15.63	6.26	.08	.79	.30
AAAA	21.69	-15.17	6.52	.35	1.09	.75

ALL PENTOSES - A+S COORDINATES

(1) Fixed-fixed interactions.
(2) Fixed-float and float-float interactions.
(3) Total
(4) Same, adjusted to have zero mean
(5) Angyal values
(6) Rao undistorted values (see section 3.3).

Table 15

Calculated energies in kcal/mole.
Best fit for ideal coordinates : 0.8 torsion

CONF	(1) FIX	(2) FLOAT	(3) TOTAL	(4) Ø MN	(5) AVG	(6) RAD
EEEE	19.49	-13.60	5.89	-.18	-1.21	-1.11
AEEE	19.96	-14.28	5.68	-.39	-.86	-.61
EAEE	21.28	-14.95	6.33	.26	-.31	-.11
EEAE	20.41	-14.10	6.31	.24	-.31	-.58
EEEA	21.79	-15.69	6.10	.03	-.76	-.43
AAEE	23.37	-17.26	6.11	.04	-.76	-.18
AEEA	21.08	-15.17	5.91	-.16	-.41	.02
EAAE	19.09	-12.62	6.47	.40	.09	-1.02
EEAA	19.59	-13.79	5.80	-.27	-.21	-.69
AEAE	23.95	-17.58	6.37	.30	.64	.80
EAEA	26.50	-20.01	6.49	.42	.74	1.19
AAAE	20.21	-14.02	6.19	.12	.39	.45
AAEA	19.33	-13.96	5.37	-.70	.29	.70
AEAA	21.96	-16.09	5.87	-.20	.74	.57
EAAA	21.18	-15.14	6.04	-.03	.79	.30
AAAA	21.13	-14.88	6.25	.18	1.09	.75

ALL PENTOSES - IDEAL COORDINATES

Columns as in table 14

Table 16

Standard deviation from Angyal values - all pentoses.

Set		(kcal/mole)
A & S	(table 1)	0.52
Ideal	(table 2)	0.66
Rao -	no angle strain	0.37
-	with angle strain	0.44

(Section 3.3)

The results are shown in tables 14 and 15, and the standard deviations from the Angyal figures in Table 16: each of these shows all 16 pentoses.

The inability of the present method to deal with the straining is shown by the large increase in σ from about 0.3 to 0.5 or more kcal/mole, which is not unexpected in the absence of any method for relaxing this strain. The Rao results, on the other hand, give about as good agreement as for the unstrained case, although again the calculation with distortion is slightly worse than that without.

3.7 DISCUSSION AND CONCLUSIONS

The present method is clearly by no means perfect. The agreement with Angyal even for the unstrained isomers is poorer than can be accounted for by statistical errors. Predicted differences between pairs such as EEEE and AEEE are, in several cases, of the wrong sign as well as differing in magnitude.

Three possible sources of this disagreement can be considered. Firstly, as previously noted, the comparison is between free energies and enthalpies. The calculation of conformational entropies has been attempted by others from the relationship

$$S = R\,(\ln q + d \ln q / d \ln T) \qquad (1)$$

where S is the relative entropy, R the gas constant, T the

temperature and q the partition function as already defined. Brant et al [98] have done this for peptides, as previously noted, and Rao [103] has made a very rough estimate for carbohydrates by assuming the only accessible states to be those defined by staggered conformations of the pendant groups, with the exception of the one in the axial configuration where the OH bond is directed across the ring. Thus in pentoses $q = q_1q_2q_3q_4$ where $q_i = 3$ for equatorial O(i) or 2 for axial, and the derivative term is taken as zero so that his $S = R \ln q$.

In a potential function such as ours, where these torsion minima are usually close to total potential minima, this approach might have a certain validity, but clearly q is a function of temperature (because more states are available at higher temperatures), and states otherthan the energy minimum are occupied, so that a much more rigorous treatment is called for.

Secondly, the present potential function might be further optimised. As has already been noted, the polar term might benefit from a varying dielectric constant.

Lastly, it must be appreciated that Angyal's values may involve some inexactitude, particularly when the method is applied to isomers for which there has been no experimental verification.

Inspection of the results does not immediately

show any trends which would suggest any of these sources as more likely than the others nor, indeed, would this be expected, since their effect on the final energies is, in most cases, complex. Inclusion of Rao's entropy terms, which although poor, may be of the correct order, does not appreciably improve agreement.

Consideration of the anomeric effect, one of the better documented aspects of the problem, shows that it emerges from the calculations, but in most cases is overestimated: thus AEEE is predicted to be stabler than EEEE contrary to Angyal, and so on. The 'Δ2 effect' (which is that the stabilisation due to the anomeric effect is enhanced for axial O(2)) is also shown: thus the EAEE - AAEE change is even greater than the EEEE - AEEE one. The origin of the anomeric effect has been ascribed [124] to the unfavourable interaction of the equatorial O(1) with a postulated 'rabbit-ear' lone pair on O(5). More recent work [125], however, has shown that the existence of such a sharply defined electron distribution is unlikely. The present conclusions thus agree with this, in that an anomeric effect is shown even when the charge on O(5) is considered to be spherically symmetrical. Since the magnitude of the effect varies with the solvent, the discrepancy in the predicted values may be an effect of the imprecise treatment given to solvent interactions, and in any case, the entropic unfavourability of the axial

configuration will tend to bring the values closer together.

In conclusion, the present work uses an accepted potential function, considering only the torsion barrier as an adjustable parameter, and yields results showing as good correlation with experiment as other methods, such as that of Rao which allows van der Waals parameters and the dielectric constant to be variable and omits entirely a torsional potential.

CHAPTER 4

THE CONFORMATIONAL ENERGY OF DISACCHARIDES

4.1 INTRODUCTION

The study of disaccharides differs from that of monosaccharides in involving the problem of linkage conformation. Linkage conformation, unlike pendant group conformation, is amenable to study by x-ray crystallography, so that results may be directly compared.

The molecules which have been studied are maltose and cellobiose, each consisting of two glucose residues linked respectively α-1,4 and β-1,4. They are the dimers corresponding respectively to amylose and cellulose.

The problem here is more difficult than with the monosaccharides, primarily because of the larger number of conformation angles. On the other hand, many linkage conformations can be rapidly eliminated by impossibly close contacts by the fixed atoms, shown by huge van der Waals energies.

Using fixed and floating as before, and (m) and (s) to refer respectively to mutual (inter-residue) and self (intra-residue) interactions,

(a) the fixed-fixed (s) terms are independent of linkage conformation and may be neglected,

(b) the fixed-fixed (m) terms are independent of pendant group conformation, but vary with ϕ/ψ, and

(c) the rest, fixed-float (s), fixed-float (m),

float-float (s) and float-float (m) vary with pendant group conformation and must, as before, be averaged.

In order to economise on computing, (c) was omitted if (b) exceeded either a preset absolute limit, or a preset amount greater than the lowest (b) + (c) found already. By inspection of the results, 500 and 50 kcals/mole respectively were fixed as these limits.

The averaging of the (c) terms was done as a plain Monte-Carlo and for practical reasons only 500 cycles were taken. This gave a much less accurate result than that in the monosaccharide case. Seven components of this averaged energy were separately recorded; these were (i) fixed-float (m) and float-float (m) van der Waals, and (ii) polar; (iii) fixed-float (s) and float-float (s) van der Waals and (iv) polar; (v) torsion about CO bonds, (vi) torsion about CC bonds and (vii) hydrogen bonding. If the jth evaluations of these are ϵ_{1j} to ϵ_{7j}, and $\epsilon_j = \sum \epsilon_{ij}$, then each was averaged by

$$\hat{\epsilon}_i = \sum_j \epsilon_{ij} \exp(-\epsilon_i/kT) \qquad (1)$$

as before, so that the correct $\hat{\epsilon}$ is the sum of the $\hat{\epsilon}_i$.

Finally, these were added to a further three terms from (b), being (viii) fixed-fixed van der Waals and (ix) polar, and (x) glycosidic bond torsion, and each term and the sum printed as a map with axes ϕ and ψ. The relative

probabilities of occupancy of each ϕ/ψ combination were also calculated in the usual way.

The summation of terms was carried out such that the hydrogen bond term included only the O'H interaction: thus it tends to be large and negative, being offset by the large positive O'O polar term. This was for practical reasons, since the latter term may occur as part of two possible hydrogen bonds if both oxygen atoms have attached hydrogens.

The calculations were performed for ϕ and ψ steps of 15° from 0° to 345°. The coordinates were initially oriented so that the glycosidic oxygen, O(4)R or O(1)N was at the origin, the R residue was arranged with C(4) on the +x semi-axis and C(1) in the +x, +y quarter plane, and the N residue with C(1) on the -x semi-axis and C(4) in the -x, +y quarter plane. Thus if P_i (x_i, y_i, z_i) is any N-residue atom and α the glycosidic angle C(4)R-O(gl)-C(1)N, the conformation given by ϕ and ψ is obtained by transforming the N-residue by

i) a rotation about the x-axis of ϕ,

ii) a rotation about the z-axis of $(\pi - \alpha)$, and

iii) a rotation about the x-axis of ψ (see fig. 24).

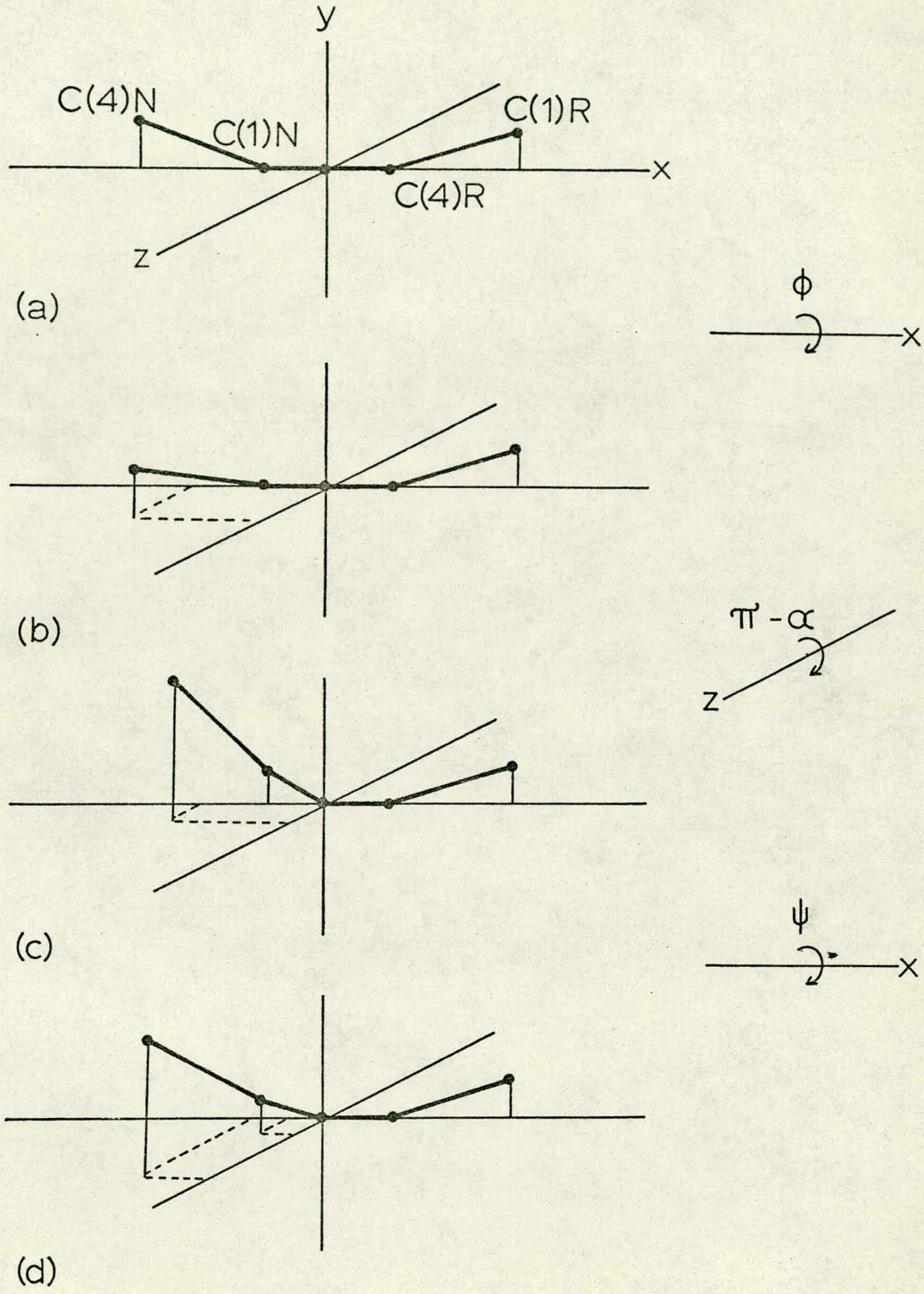


Fig. 24. Generation of disaccharide coordinates

This is achieved (Appendix 3) by:

$$\underline{P}_i' = \underline{\underline{T}} \cdot \underline{P}_i \qquad (2)$$

where

$$\underline{\underline{T}} = \begin{bmatrix} -\cos\alpha & \sin\alpha\cos\phi & \sin\alpha\sin\phi \\ -\cos\psi\sin\alpha & -\cos\alpha\cos\psi\cos\phi - \sin\psi\sin\phi & -\cos\alpha\cos\psi\sin\phi + \cos\phi\sin\psi \\ \sin\psi\sin\alpha & \sin\psi\cos\alpha\cos\phi - \cos\psi\sin\phi & \sin\psi\cos\alpha\sin\phi + \cos\psi\cos\phi \end{bmatrix}$$

The coordinates used were Arnott and Scott (see section 3.4) with a glycosidic angle of 117°.

4.2 CALCULATION OF HELIX PARAMETERS

If the linkage conformation between each pair of residues in a homopolysaccharide is the same, then the polysaccharide will be helical. Such helical structures have been observed (section 1.5) and characterised in terms of h, the projected height of a residue on the helix axis, and n, the number of residues per helix turn (see fig. 9).

Miyazawa [126] has developed a method for calculating n and h from ϕ to ψ, as follows. Consider each residue to be composed of three 'bonds' between P_1 (O(1)), P_2 (C(1)), P_3 (C(4)), and P_1' (O(4)) (fig. 25). Let r_{ij} be the distance P_iP_j, μ_i the angle $P_{i-1}P_i\,P_{i+1}$, and t_{ij} the dihedral angle between planes $P_{i-1}\,P_i\,P_j$ and $P_i\,P_j\,P_{j+1}$.

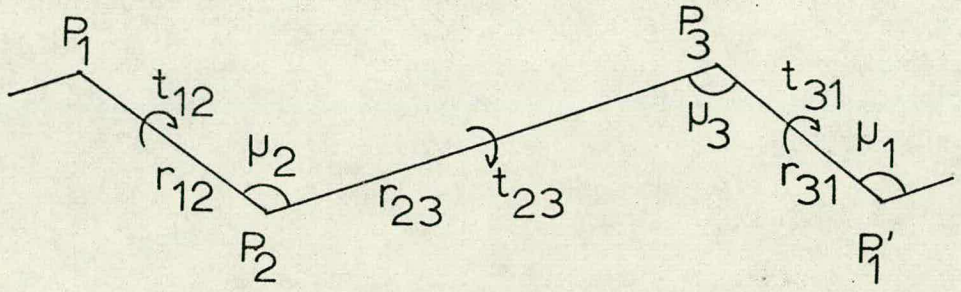


Fig. 25. Internal coordinates for helix parameter calculation; P_1 and P_1' represent glycosidic oxygens, P_2 and P_3 are C(1) and C(4) respectively.

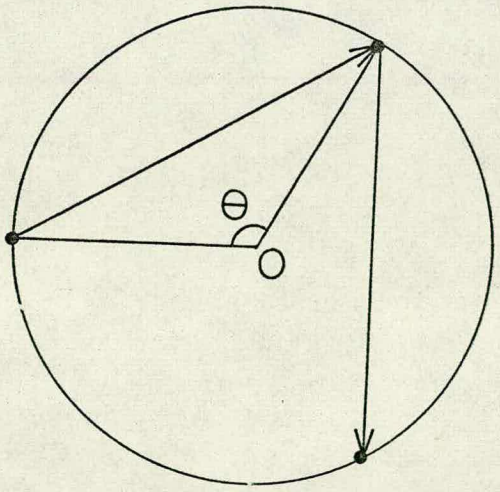


Fig. 26. Projection down the helix axis, O, showing θ. The arrows represent residues, the black dots, successive glycosidic oxygens.

Then,

$$\begin{aligned}\alpha=\cos\tfrac{1}{2}\theta = {} & \cos(\tfrac{1}{2}(+t_{12}+t_{23}+t_{31}))\sin\tfrac{1}{2}\mu_1 \sin\tfrac{1}{2}\mu_2 \sin\tfrac{1}{2}\mu_3 \\ & -\cos(\tfrac{1}{2}(-t_{12}+t_{23}+t_{31}))\cos\tfrac{1}{2}\mu_1 \sin\tfrac{1}{2}\mu_2 \sin\tfrac{1}{2}\mu_3 \\ & -\cos(\tfrac{1}{2}(+t_{12}-t_{23}+t_{31}))\sin\tfrac{1}{2}\mu_1 \cos\tfrac{1}{2}\mu_2 \sin\tfrac{1}{2}\mu_3 \\ & -\cos(\tfrac{1}{2}(+t_{12}+t_{23}-t_{31}))\sin\tfrac{1}{2}\mu_1 \sin\tfrac{1}{2}\mu_2 \cos\tfrac{1}{2}\mu_3 \end{aligned} \qquad (1)$$

and

$$\begin{aligned}\beta = h\sin\tfrac{1}{2}\theta = {} & (+r_{12}+r_{23}+r_{31})\sin(+\tfrac{1}{2}t_{12}+\tfrac{1}{2}t_{23}+\tfrac{1}{2}t_{31}) \\ & \sin\tfrac{1}{2}\mu_1 \sin\tfrac{1}{2}\mu_2 \sin\tfrac{1}{2}\mu_3 \\ & -(-r_{12}+r_{23}+r_{31})\sin(-\tfrac{1}{2}t_{12}+\tfrac{1}{2}t_{23}+\tfrac{1}{2}t_{31}) \\ & \cos\tfrac{1}{2}\mu_1 \sin\tfrac{1}{2}\mu_2 \sin\tfrac{1}{2}\mu_3 \\ & -(+r_{12}-r_{23}+r_{31})\sin(+\tfrac{1}{2}t_{12}-\tfrac{1}{2}t_{23}+\tfrac{1}{2}t_{31}) \\ & \sin\tfrac{1}{2}\mu_1 \cos\tfrac{1}{2}\mu_2 \sin\tfrac{1}{2}\mu_3 \\ & -(+r_{12}+r_{23}-r_{31})\sin(+\tfrac{1}{2}t_{12}+\tfrac{1}{2}t_{23}-\tfrac{1}{2}t_{31}) \\ & \sin\tfrac{1}{2}\mu_1 \sin\tfrac{1}{2}\mu_2 \cos\tfrac{1}{2}\mu_3 \end{aligned} \qquad (2)$$

where θ is angle subtended at the helix axis by one residue, when projected down the axis (see fig. 26), which is $2\pi/n$. Thus,

$$n = \pi \ / \text{arc cos } \alpha \qquad (3)$$

and

$$h = \beta \ / \ \sqrt{1-\alpha^2} \qquad (4)$$

Unfortunately, several sets of conventions have been applied to the helix parameters. Clearly, if $n < 2$, the helix can equally be considered as of the opposite sense with $\theta' = 2\pi - \theta$ and $n' = n/n-1$. Helices of opposite senses (left or right handed) may be formally denoted by positive

and negative n or h. Following Rees [127] in this work n will be taken as positive and negative for right- and left-handed helices respectively and $/n/ \geq 2$, and h always positive.

These may be generated from the above α and β (eqns. 1 and 2) thus:

$$h = \left| \beta / \sqrt{1-\alpha^2} \right| \qquad (5)$$

$$n = \text{sign}(\alpha\beta) \cdot \pi / \arccos |\alpha| \qquad (6)$$

In this convention, $n = x$ is equivalent to $n = -x$ if $x = 2$ or $h = 0$.

For geometrical details (of dihedral angle calculation etc.) see Appendix 3.

The method detailed gives n and h as functions of ϕ and ψ. Occasionally, however, the values of ϕ and ψ corresponding to a given n and h are required. These can be found approximately from graphical interpolation, but for many purposes, such as the calculation of coordinates for lengths of polysaccharide, more accurate values are desirable.

To achieve this, a program was written to do a least-squares fit of ϕ and ψ to any given n and h. This used the simplex procedure outlined in Appendix 1. In order to avoid confusion between the several points with the same n and h values, a small (1°) steplength was chosen.

The procedure then converged to less than .01° in both ϕ and ψ usually after fewer than 200 evaluations, when started from within about 10° in both angles.

This method was used to fix the positions of possible helical structures on the ϕ/ψ maps in the next sections.

4.3 MALTOSE: THE α-1,4 LINKAGE - RESULTS

The results of the calculation described in section 4.1 on maltose are shown in fig. 27. The contours are 5 and 10 kcal/mole above the minimum energy conformation shown by M. A subsidiary calculation with smaller steplength (2°) in ϕ and ψ was carried out in order better to fix this minimum both in position and energy; this showed it to lie in fact very close to the originally evaluated position.

The relative probabilities of occupancy of each of the areas of less than 5 kcal/mole were calculated and are shown in table 17, from which it can be seen that the largest area is greatly favoured.

The other points marked on fig. 27 show crystal structures (C) and the established helical V form of amylose when complexed with iodine (V). Also shown (H) are the calculated linkage conformations in the cyclohexa-amylose-water complex studied by Saenger[128,129] showing,

Fig. 27. Potential energy map for maltose (over)

Contours are 5 and 10 kcal/mole above the minimum.

The symbols indicate:

M - the minimum of the potential

C - crystal structures:
upper left - β-maltose [131]
upper right - cyclohexaamylose [132]
lower - methyl β-maltoside [133]

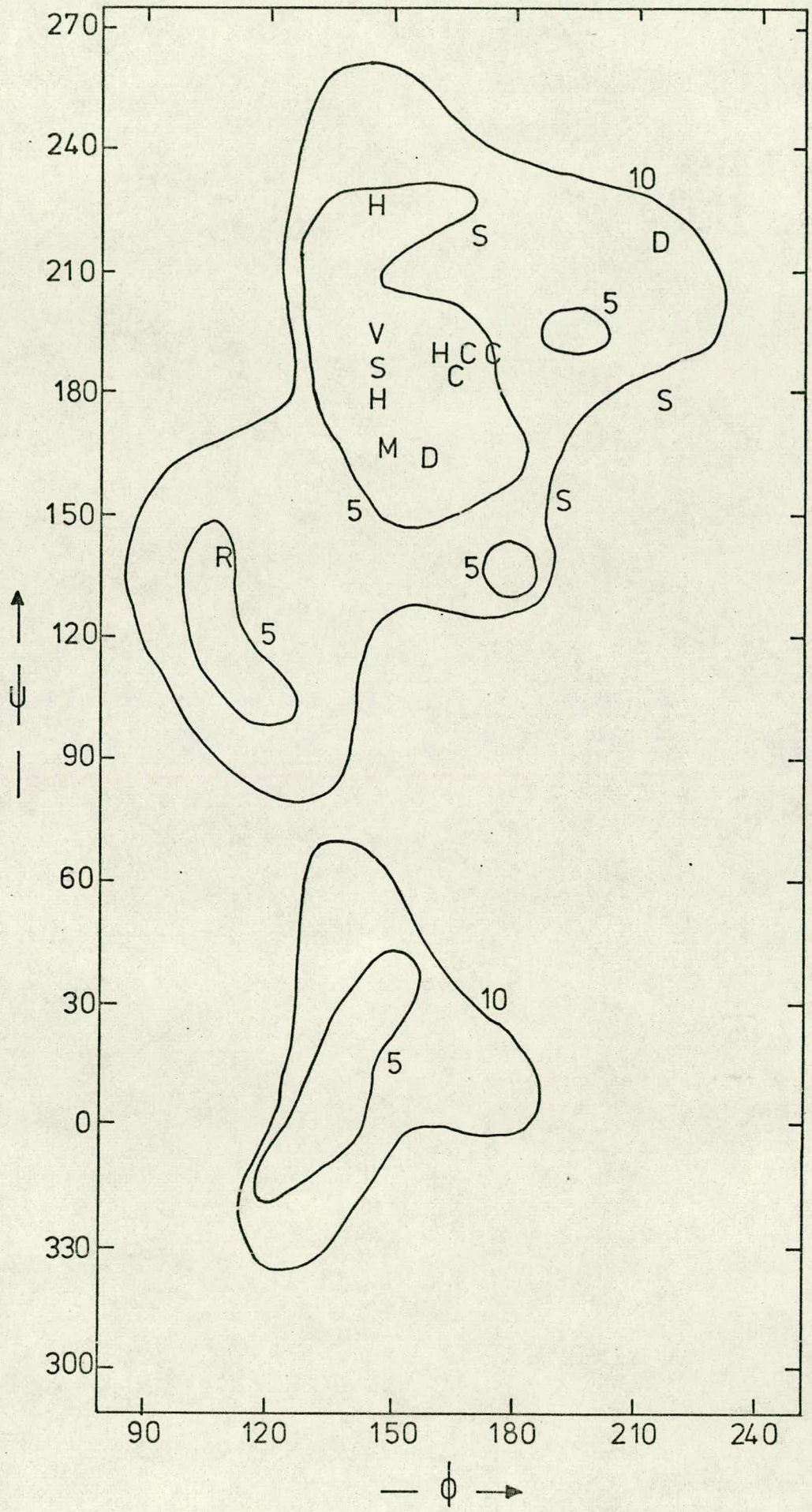
H - Saenger cyclohexaamylose (see text)

V - V-amylose, $n = -6$, $h = 1.32$ Å

S - B-amylose, single helix model, $n = 6$, $h = 1.73$ Å: top and right, right-handed helix; bottom and left, left-handed helix.

D - B-amylose, double helix model, best approach: $n = 6$, $h \doteq 3$ Å: upper, right-handed helix; lower, left-handed helix

R - solution conformation predicted by Thom [54] from optical rotation work.


270
240
210
180
150
120
90
60
30
0
330
300
90
120
150
180
210
240
ψ
φ
10
H
S
D
5
V
S
H
H C C
C
M
D
S
5
S
R
5
5
10
5

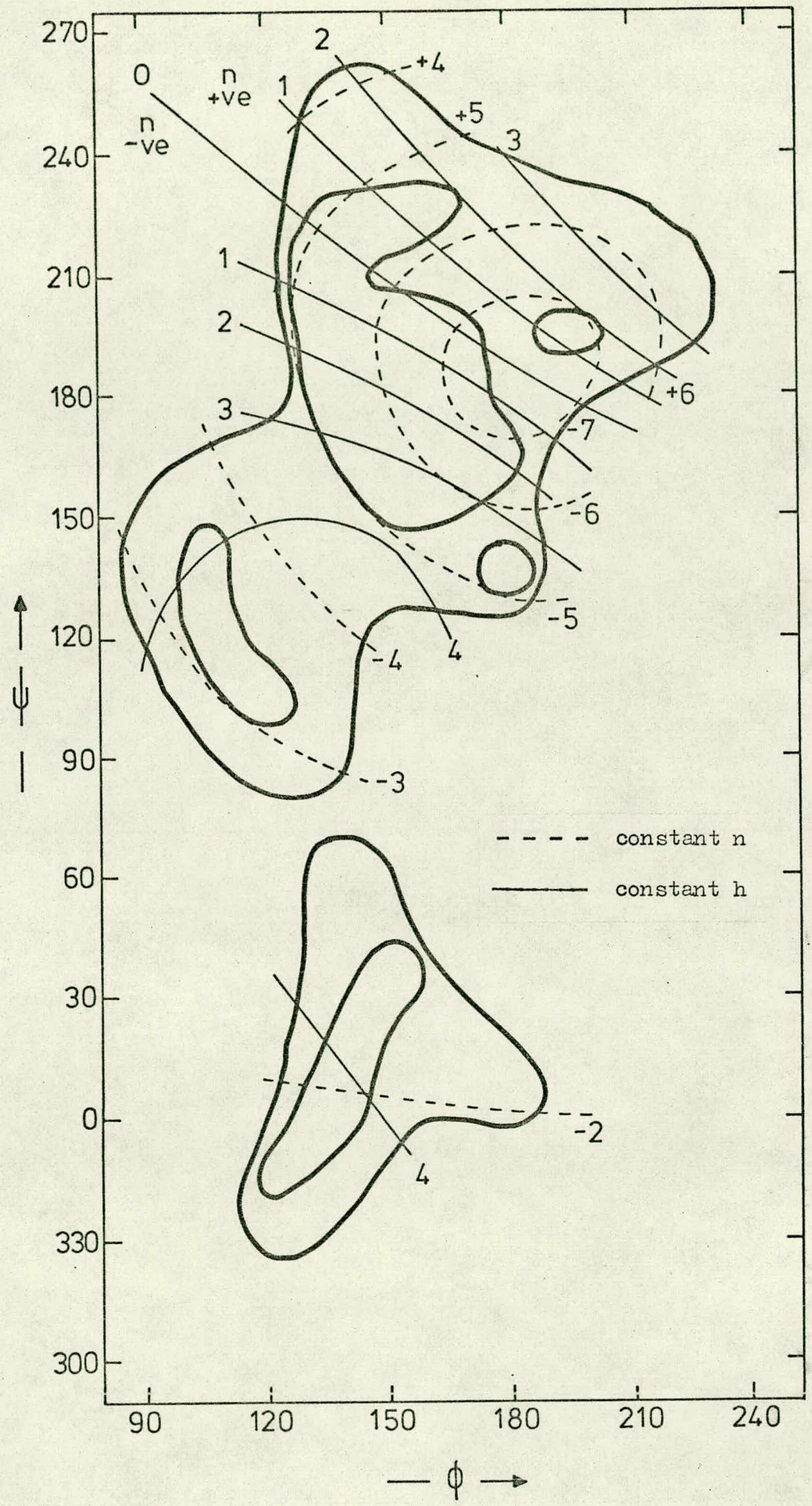


Fig. 28. Helix parameters on the energy contours for maltose.

TABLE 17

Thermodynamic probabilities of the low energy zones at 25°C

Zone around ϕ	Zone around ψ (deg.)	Thermodynamic probability (percentage)
150	180	94.5
180	130	3.3
105	120	1.1
130	0	0.8
195	195	0.2

close to the crystal values, the mean of four similar linkages, and, separately, the two linkages to the 'twisted' sixth residue.

Possible conformations of B amylose, the other important form, are also shown. From diffraction studies this is known to have $|n| = 6$ and a fibre repeat of 1.73 Å. Suggested structures are thus a single helix (s) with $h = 1.73$ Å or a double helix with $h = 3.46$ Å.

From helix parameter calculations (see fig. 28), this latter structure is shown to be impossible with A & S geometry, since with $n = -6$, the maximum possible h is just under 3.0 Å, so that the double helix could form only with some distortion of the rings. The 'best approach' to this structure is shown (D), and the question is further discussed in section 5.5.

Optical rotation studies in aqueous solution and the application of Rees' rule (see section 1.7) have led Thom [54] to favour the conformation shown by R. Although this zone is of low probability, it can easily be seen from models that it lacks the stabilising O(2)N to O(3)R hydrogen bond present in the region which is assigned high probability by the present calculation. In aqueous solution this hydrogen bond will be much less likely because of competitive hydrogen bonding with water molecules, so that the zone containing R will be relatively more favourable.

However for methyl β-maltoside dissolved in dioxan, a solvent which cannot form hydrogen bonds, Thom[54] reports an observed linkage rotation of -23° at 25°C, a value which Rees' rule predicts for certain conformations within 5° in ϕ and ψ of the calculated energy minimum.

Examination of the terms which contribute to the total energy gives information as to the predominant forces within the molecule.

The shape of the low energy regions is determined primarily by the fixed atom van der Waals interactions, which explains the success of the hard sphere approach. Within these regions, however, both fixed and floating atom polar terms exert a marked influence, the latter being complicated by the hydrogen bond potential which may offset any large O-O repulsions.

Torsion terms generally average out relatively constant as do the other intraresidue terms, as would be expected, since their dependence on ϕ and ψ is a secondary effect.

A comparison of the results calculated here with several reported in the recent literature is shown in fig. 29, which shows the 5 kcal/mole contour for each. The comparison is approximate since other workers have not all reported the depth of the minimum and thence the

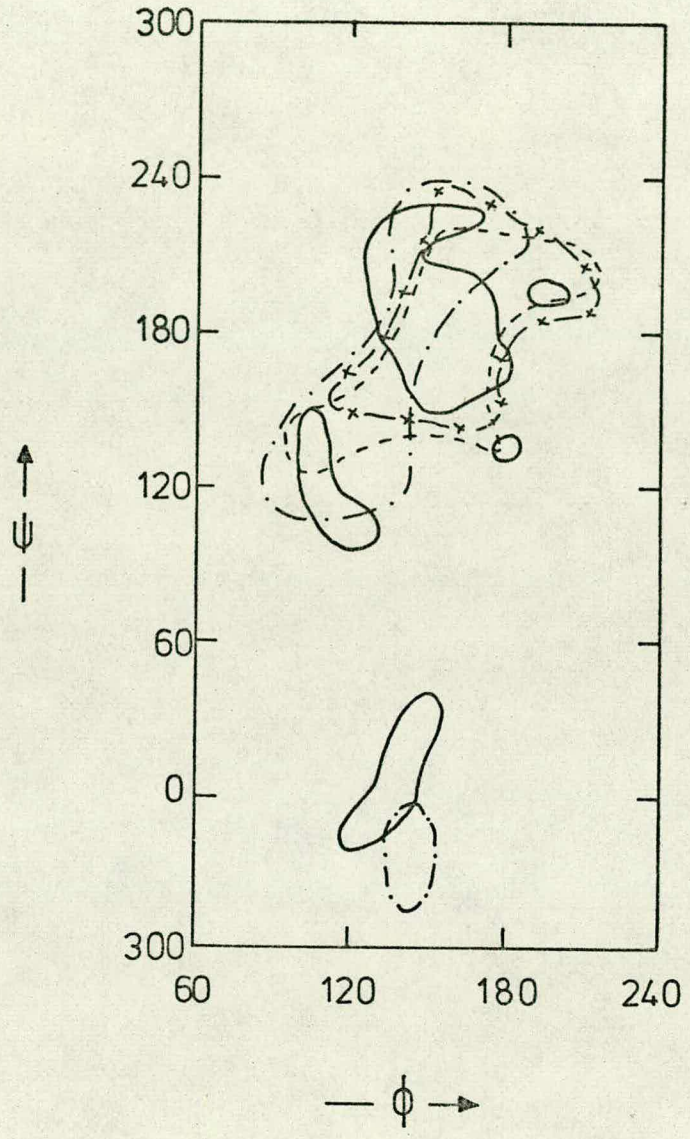


Fig. 29. Comparison of potential energy calculations for maltose. Contours are ~5 kcal/mole above the minimum.

————	present calculation
- - - - -	Rao (Flory potential) Brant and Dimpfl is very similar
- x - x	Rao (Kitaigorodsky potential) Blackwell et al is very similar
-.-.-.-.-.	Pullman and co-workers

See text for details and references.

absolute heights of their contours, and also because several use ϕ/ψ conventions based on coordinate-dependent factors so that the positions given may be up to about 5^{o} out.

The other studies are as follows:

i) Rao and co-workers[78] using the Flory potential (6-exp) with no polar or torsion terms,

ii) Rao and co-workers [73] using the Kitaigorodsky potential with no polar or torsion terms,

iii) Blackwell, Sarko and Marchessault [25], similar to (i) but including a H-bond term,

iv) Pullman and co-workers [95] using a quantum mechanical treatment.

v) Brant and Dimpfl [130] using a Lennard-Jones (6-12) potential and some polar terms but no torsion terms.

It can be seen that the various methods give essentially similar information, although no treatment other than the present one accounts for both the crystal and solution conformations. It is also clear that the inclusion of a full polar and hydrogen bond treatment substantially modifies the map, giving in general sharper minima and favouring areas of lower ϕ more than the simpler methods.

This calculation, therefore, is consistent with

the known conformations of the α-1,4 linkage in the solid state and in aqueous solution, whilst, unlike previous work in the field, taking account of the positions of all the atoms in the molecule considered, and utilising a potential function which follows more closely the known properties of the molecule.

4.4 CELLOBIOSE: THE β-1,4 LINKAGE - RESULTS

Calculations similar to those performed on maltose were repeated for cellobiose: the results are shown in fig. 30 with helix parameters in fig. 31. The relative probabilities of occupancy of the regions of less than 5 kcal/mole are shown in table 18.

On fig. 30 are shown several crystal structures (C) and the generally accepted [65] bent-chain conformation of cellulose which has n = 2 and h = 5.15 Å. Optical rotation studies [48] suggest that the conformation of cellobiose in aqueous solution is close to its crystal conformation. All of these are close to the calculated minimum, and derive much of their stability from a O(5)N to O(3)R hydrogen bond.

The region around $\phi = 0^{\circ}$, $\psi = 180^{\circ}$ is similarly stabilised by hydrogen bonding between O(6)R and O(5)N or O(6)N. Although this is a possible structure for cellobiose, its low h value is such as to exclude any possibility of

Fig. 30. Potential energy map for cellobiose.

Contours are 5 and 10 kcal/mole above the minimum.

The symbols indicate:

M - the minimum of the potential

C - crystal structures;

upper - β-cellobiose [136]

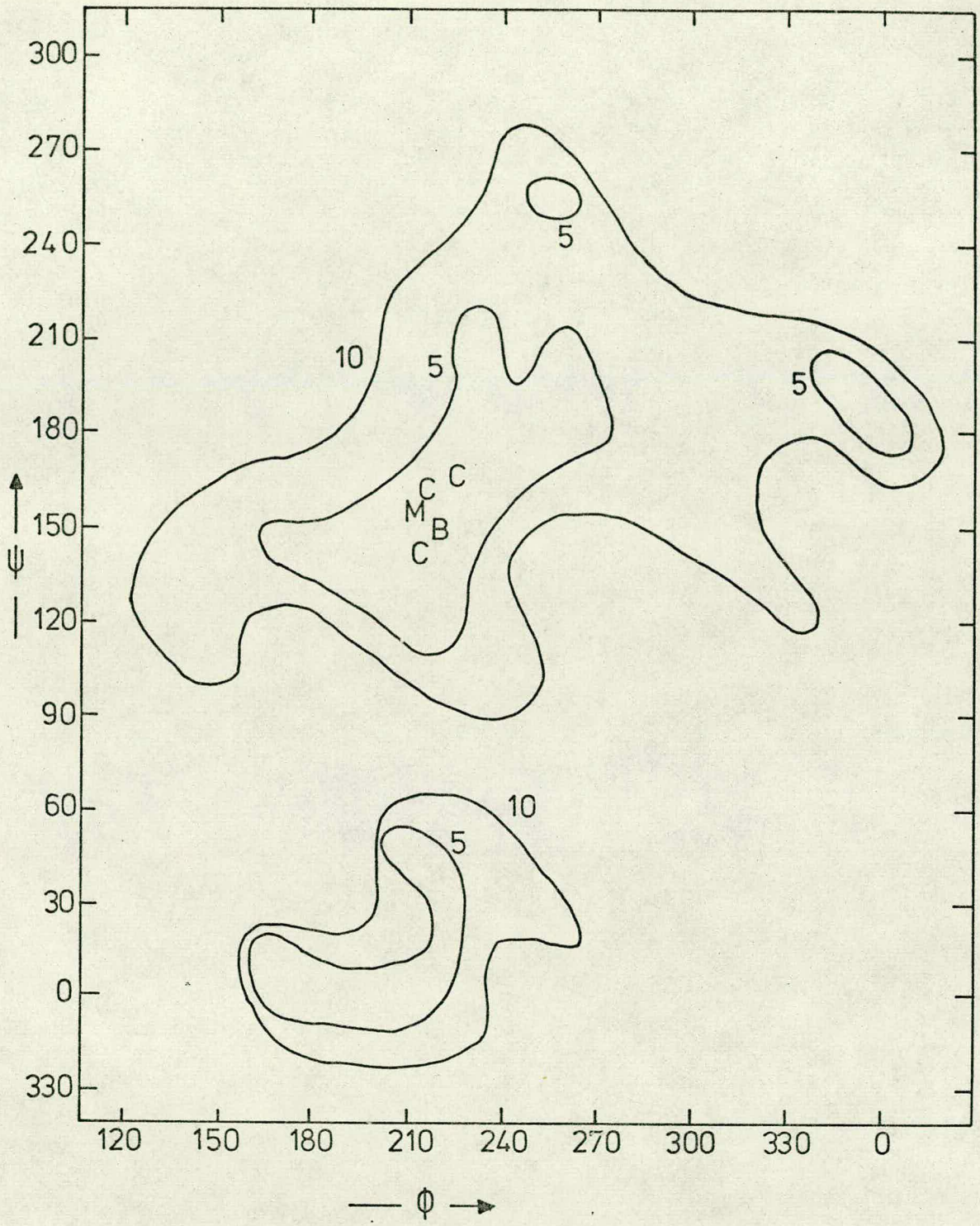
(this is also its aqueous solution conformation)

middle - α - lactose [137]

lower - methyl β-cellobioside [138]

B - the bent chain conformation of cellulose,

n = 2, h = 5.15 Å.


300
270
240
210
180
150
120
90
60
30
0
330
ψ
120
150
180
210
240
270
300
330
0
φ
10
5
5
5
10
5
M
C
C
C
B
C

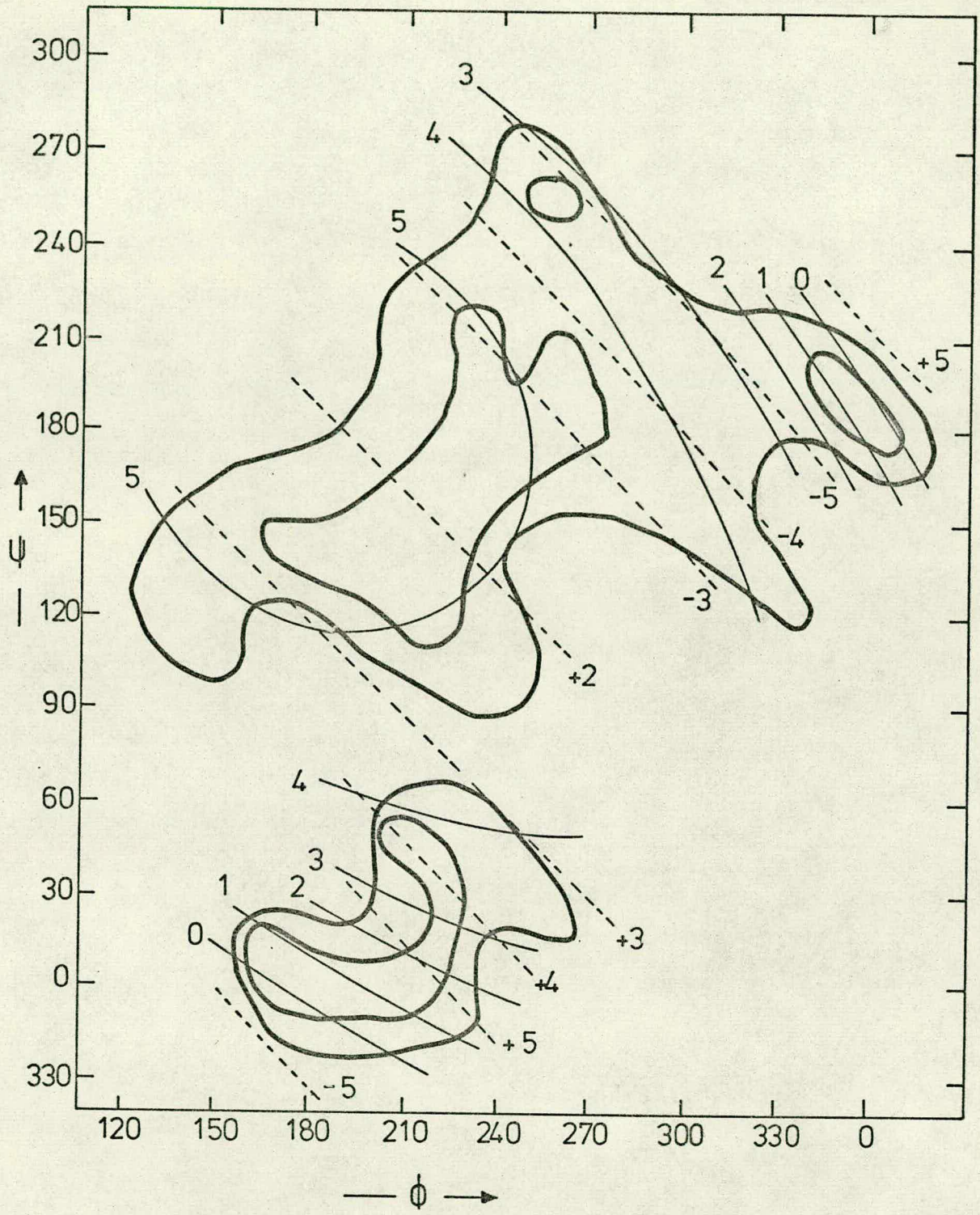


Fig. 31. Helix parameters on the energy contours for cellobiose.

- - - constant n ——— constant h

TABLE 18

Thermodynamic probabilities of the low energy zones at 25°C.

Zone around ϕ	ψ (deg.)	Thermodynamic probability (percentage)
210	150	77.0
210	0	19.8
260	260	1.6
0	180	1.6

its regular repetition in the polymer. The region around $\phi = 210^{o}$, $\psi = 0^{o}$, corresponding to the formation of an O(6)N - O(6)R hydrogen bond, is similarly affected.

Examination of the contributions to the total energy map yields similar conclusions to those for the maltose calculation. However, since the β-linkage is less constricted in general than the α, the fixed atom van der Waals terms permit many more linkage conformations, and the other terms are correspondingly more important within these regions. The correlation between low energy zones and specific hydrogen bonds is in this way much more marked than in maltose.

Despite this apparent openness, however, the minima in maltose and cellobiose are of almost the same energy (maltose being about 0.8 kcal/mole less).

A comparison of several published energy maps with the present one is shown in fig. 32, with the same qualifications as before.

The other studies are as follows:

i) Pullman and co-workers [95] using a quantum mechanical treatment.

ii) Yathindra and Rao [134] using a Kitaigorodsky function and glycosidic bond torsion but no polar terms.

iii) Sathyanayarana and Rao [135] using a similar function

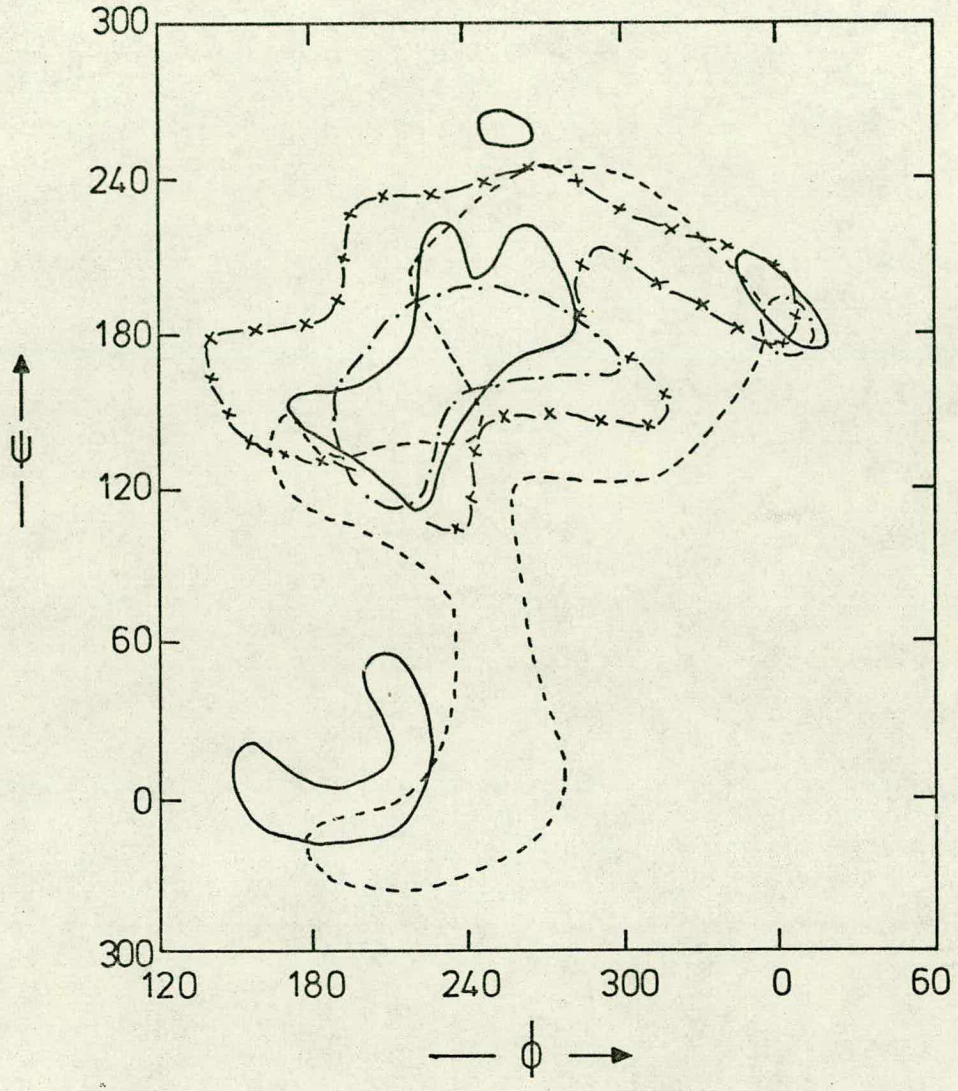


Fig. 32. Comparison of potential energy calculations for cellobiose. Contours are ~5 kcal/mole above the minimum.

——— present calculation
- - - - Pullman and co-workers
-.-.-.- Sathranayarana and Rao
-x-x-x- Yathindra and Rao

See text for details and references.

to (ii) but including hydrogen bonding.

In addition, Rees and Skerret have published[127] maps using Kitaigorodsky, Flory and Liquori functions evaluated only inside their hard-sphere allowed zone, so that the 5 kcal/mole contour is not available.

Perhaps the most surprising aspect of this comparison is the diversity in the various maps, which is in contrast to the maltose case discussed previously. This is presumably due mainly to the flatter energy surface of cellobiose.

As in maltose, the inclusion of polar and hydrogen bond terms clearly influences the map to a large extent, and whilst here several simpler treatments are also consistent in general with the known conformations, the present treatment yields an energy minimum very close indeed to the observed conformations.

4.5 CONCLUSIONS

The results presented in this section show that the chosen potential function matches well the observed properties of disaccharides.

The hydrogen bond potential, in particular, clearly reproduces well observed tendencies to form such bonds, without, contrary to many published calculations,

the need to preselect those oxygen atoms which are expected to partake in them.

The present method, while practicable for mono- and disaccharides, has the disadvantage of requiring a great deal of computation in its statistical averaging. For this reason, given present resources, recourse must be made to simplified treatments when polysaccharides are considered.

A number of such methods are discussed and applied in the next chapter.

CHAPTER 5

APPLICATIONS OF CONFORMATION CALCULATIONS TO POLYSACCHARIDE PROBLEMS

5.1 INTRODUCTION

Chapters 3 and 4 have dealt respectively with calculations on mono- and disaccharides. This chapter shows how this method may be generalised to certain polysaccharides.

The conformation of regular polysaccharides is, in general, governed by three factors. Probably the most important is the interaction between adjacent residues, the linkage conformation. Secondly, there may exist interactions between residues which are not adjacent on the chain, but because of the form of the polymer, are close together in space. Lastly, there is the interaction with other molecules either of different species (such as solvent interactions) or of the same species.

This last factor is obviously most important in the solid state, although if the forces are sufficiently attractive, various types of such structures persist in solutions or gels. Two types of structuring are worth distinguishing; those where the molecules are in some way intertwined, such as double or triple helical structures, and those where the molecules pack together in a more or less ordered 'crystallite' structure. Clearly both of these may be present in a system at once.

The determination of possible regular structures

of polysaccharides may therefore be approached as follows:

i) Linkage conformation calculations on appropriate disaccharides with helix parameter determinations show the likely regions for ϕ and ψ, and the corresponding n and h values. Crystal packing considerations often limit possible n values to integral values such that the molecule has screw symmetry of 2, 3, 4 or 6.

ii) Model building, either physically or by calculation, shows which of these are likely to suffer from remote contacts: in general those with too low h. It may also show whether double or triple helical structures can exist.

iii) X-ray fibre diffraction evidence, when available, gives directly n and the fibre repeat distance, which is a multiple of h.

This type of deduction may be modified or adapted to take advantage of any other facts known about the polymer, such as cation binding or hydrogen bonding.

This chapter illustrates the method as used in various polysaccharide systems.

5.2 THE CATION EGGBOX MODEL

Interactions between polymers of uronic acids and inorganic cations are important both biologically and

industrially, but little is known about the structure of the complexes thus formed.

Recent experimental work [54, 139] using circular dichroism (see section 1.7) on alginates (VIII) has shown the existence of a specific binding of calcium ions which occurs cooperatively with the sol - gel transition. Alginates are composed of D-mannuronate and L-guluronate subunits which are arranged in blocks of all mannuronate, all guluronate, and alternating mannuronate and guluronate, and it is shown that it is the guluronate blocks which bind the calcium ions.

Similar effects have also been shown [140] in pectin (IX), which is a D-galacturonate polymer. Studies by Gould and others [141] on the crystal structure of sodium calcium and sodium strontium galacturonates have shown that the divalent ion is located (see fig. 33) at the apex of a pyramid, the base of which is an O(5), O(6), O(water) triangle. Williams [142] has subsequently suggested the 'ideal' condition as that with the Ca-O(6) distance as 2.6 $\pm$ 0.2 Å and as many as possible other Ca-O distances as 2.7 $\pm$ 0.3 Å. Similar Ca-O distances have been found in complexes of calcium ions with uncharged sugars by Bugg [143].

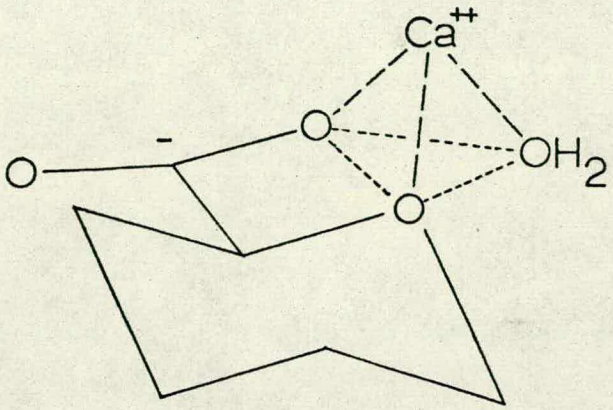


Fig. 33. Coordination triangle formed by calcium ions with uronic acids.

Calculations were performed to establish the ability of the relevant polysaccharides to form such favourable coordination complexes. The calculations were carried out on the appropriate disaccharides, which are α-L-gulosyluronate-α-L-guluronate (gul-gul), β-D-mannosyluronate-β-D-mannuronate (man-man), α-L-gulosyluronate-β-D-mannuronate (gul-man), β-D-mannosyluronate-α-L-guluronate (man-gul) and α-D-galactosyluronate-α-D-galacturonate (gal-gal).

In each case it was established whether the criteria defined above could be met for O(6) with any two of O(5) on the same residue, and O(2) and O(3) on the next

residue in either the reducing and non-reducing directions.

The coordinates taken were Arnott and Scott (see section 3.4) with the COO^- group planar and with C(5)-C(6)-O(6) and -O(7) 116.8°, and C(6)-O(6) and O(7) 1.25 Å. The dihedral angle O(5)-C(5)-C(6)-O(6) was established in various ways. From the already mentioned crystal structure determinations [141], one of strontium Δ-4,5-digalacturonate [144] (which contains four non-equivalent residues) and one of potassium glucuronate [145], a total of seven examples, the angle was calculated to fall always within the limits of -35° to +15° (positive angle being towards H(5)). The calculations were performed with the dihedral

(a) at the value optimal to formation of the triangle under consideration,

(b) stepped in 10° steps from -35° to +15°, and

(c) fixed at -5°.

It was found that the differences were in every case marginal: the quoted results are those from (c).

The calculation of the triangles was carried out for every linkage conformation taken in 10° steps allowed by the hard sphere criteria previously detailed. Coordination of the cation with the three oxygens atoms at P_1 = O(6), P_2 and P_3 was deemed to be possible if(fig. 34):

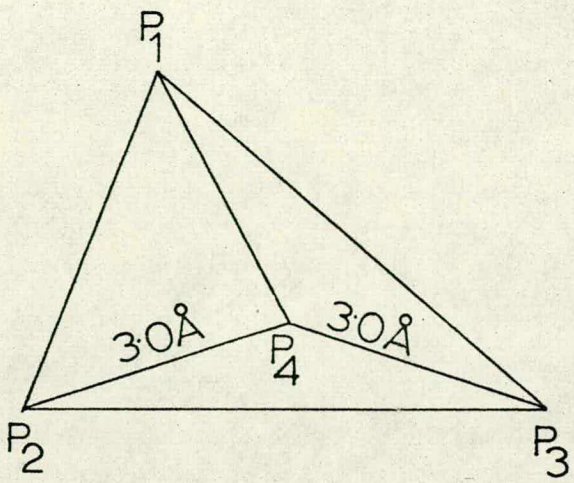


Fig. 34. Test case for cation coordination.

---

i) $P_1P_2 < 5.8$ Å, $P_1P_3 < 5.8$ Å, $P_2P_3 < 6.0$ Å, which are the limiting conditions for coordination with two oxygen atoms, and

ii) $P_4 = Ca^{++}$ existed such that $P_4P_1 < 2.8$ Å, when P_4 was in the plane of $P_1P_2P_3$, which is the limiting condition for coordination with three oxygen atoms, since any out-of-plane position for P_4 will yield a larger P_1P_4 distance.

It can be seen from models that this covers all possibilities allowed by the hard sphere restrictions on the atoms concerned.

For the three dimers corresponding to homopolysaccharides, helix parameters were calculated (see section 4.5). For man-gul and gul-man taken together, helix parameters for the alternating heteropolysaccharide were calculated.

Figs. 35 to 39 show the hard sphere maps together with the number of possible types of triangle possible for each linkage conformation. Helix parameters are superposed for the homopolymer cases.

The results may be summarised as follows.

Polyguluronate can form chains with two or threefold symmetry with good binding sites involving O(6), and in most cases O(5), with O(2) and O(3) on the next residue in the N-direction, giving probable four oxygen coordination. This is in agreement with the experimentally observed twofold structure [146].

Polygalacturonate can do so too, but mainly for only three oxygens, for fewer linkage conformations and only for n = 2. It seems therefore not unlikely that its calcium salt exists in a twofold form, although the sodium salt has been shown to be threefold [147, 148]. The O(6)-O(2) distance here is however markedly less than with polyguluronate (typically 3.4 Å against 3.8 Å) which may explain the ion selectivity difference between the two observed by

Figs. 35 to 39. Coordination possibilities for disaccharide residues.

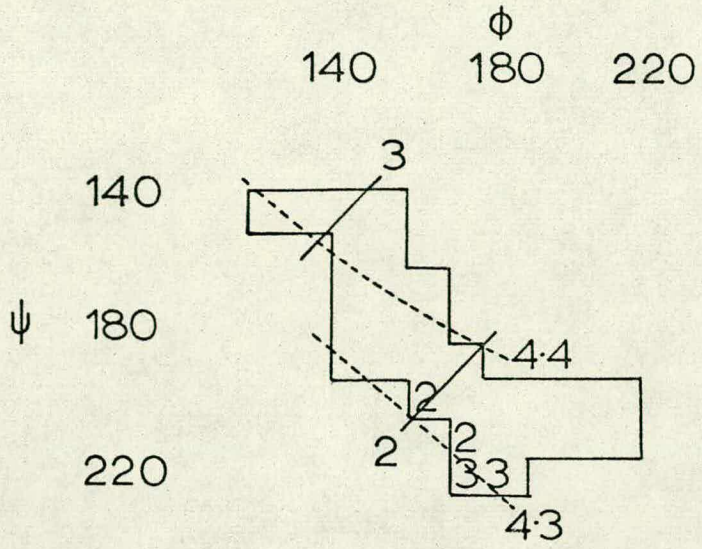
Solid outlines show the hard sphere permitted regions.

Full line contours are constant n.

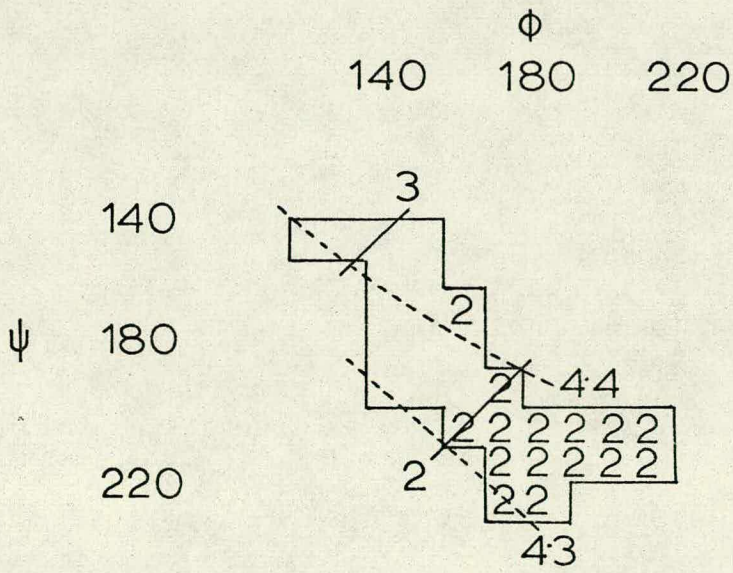
Dashed line contours are constant h in Å.

Numbers within the outline are the number of triangles possible (see text), maximum 3. Blank indicates 0 or 1.

The upper diagram in each case is from O(6) towards oxygen# atoms on the next residue in the direction towards the reducing end of the chain; the lower, towards the non-reducing end.

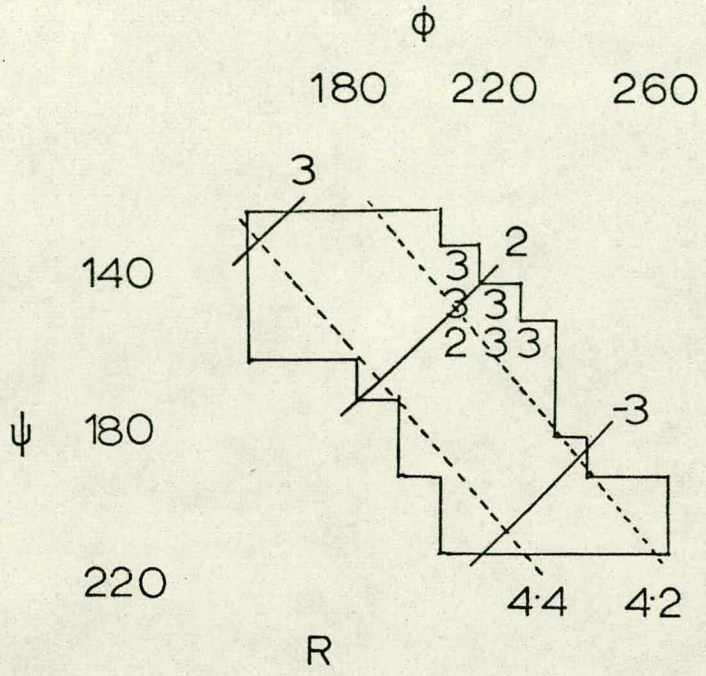


R



N

Fig 35 Gal-Gal



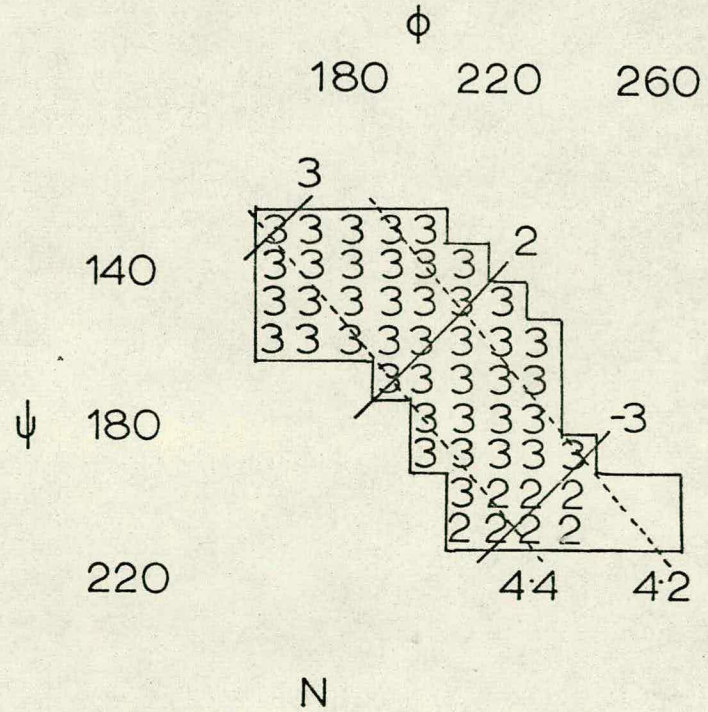
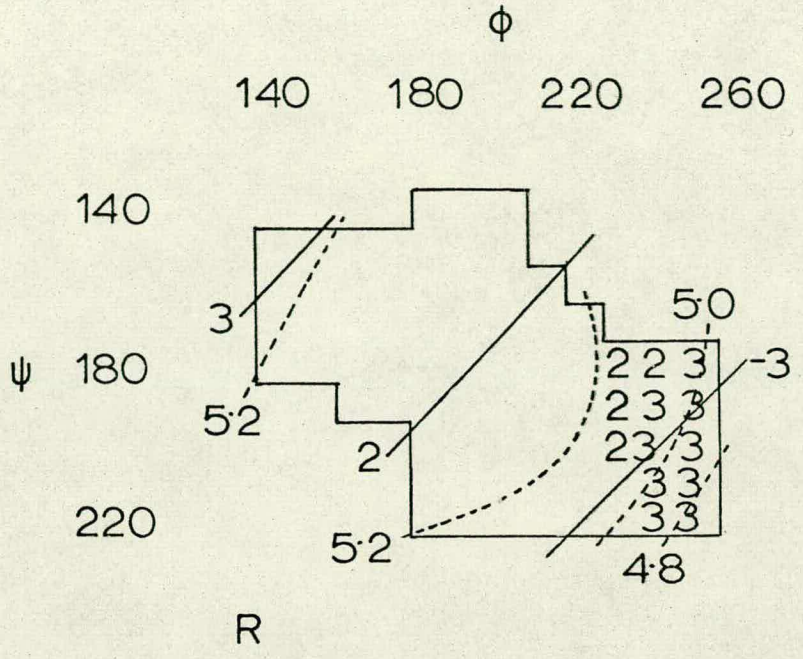
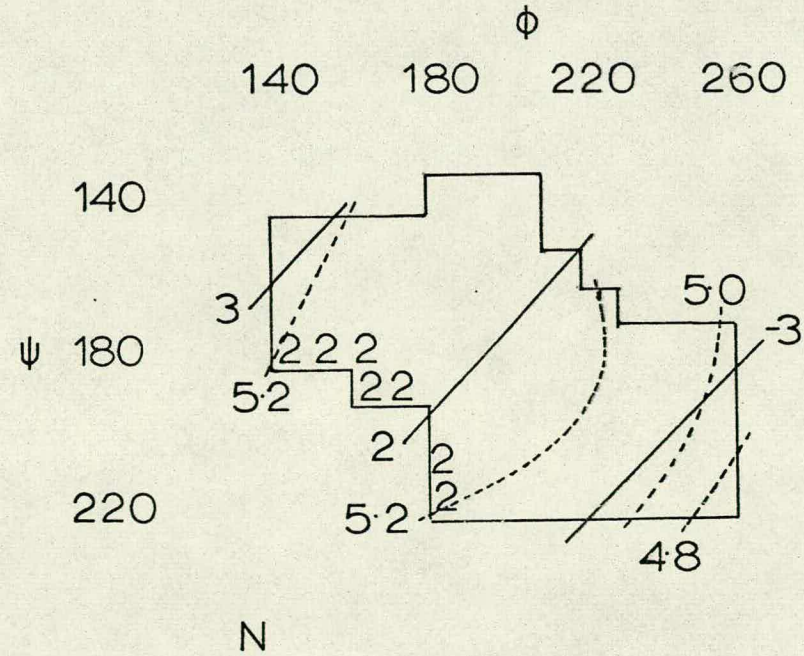


Fig 36 Gul-Gul

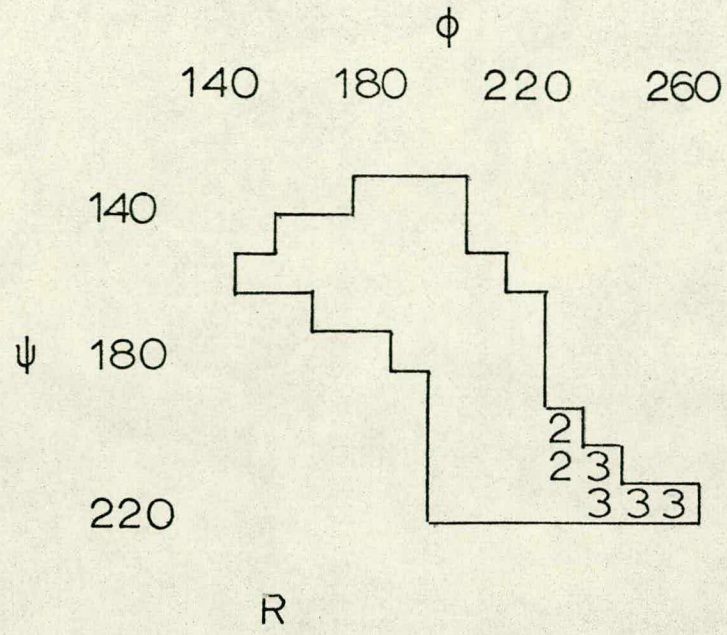


R

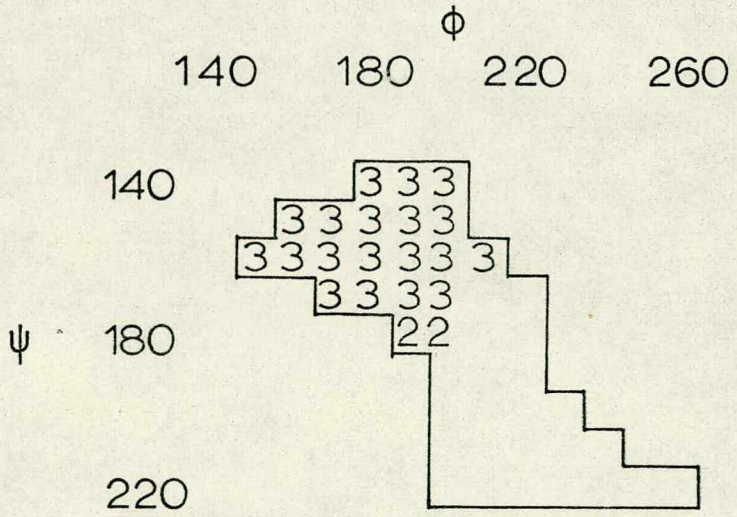


N

Fig 37 Man-Man

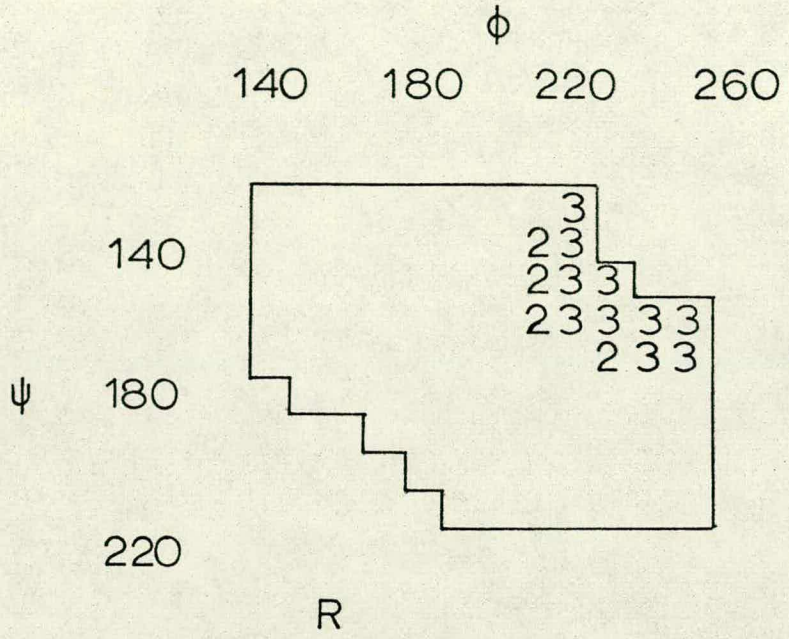


R



N

Fig 38 Gul-Man



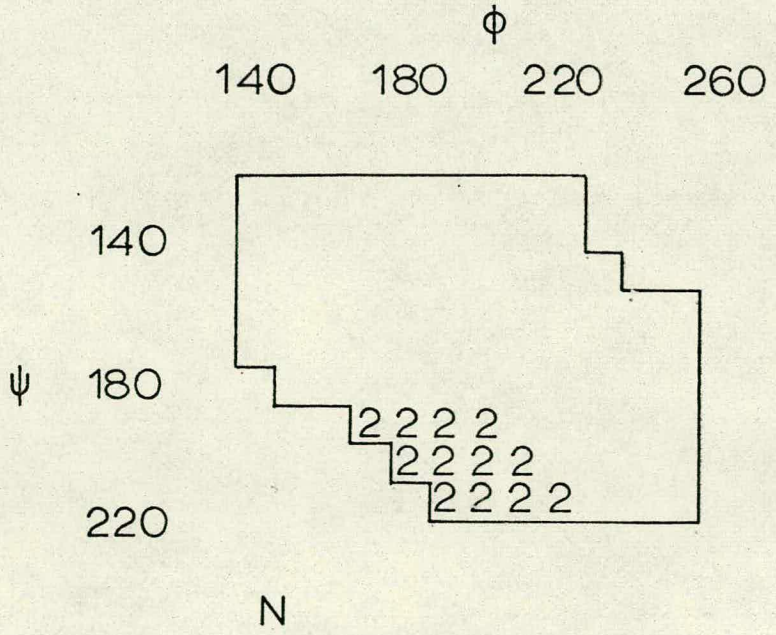


Fig 39 Man-Gul

Smidsrød and co-workers. Although both bind Ca^{++} with about the same affinity [149], polyguluronate binds the larger Sr^{++} and Ba^{++} ions much more strongly than polygalacturonate, and vice versa with the smaller Mg^{++} ions [150].

Polymannuronate chains form sites with both two and threefold symmetry, but inspection of models shows that the β-linkage combined with the observed threefold screw symmetry [146] leads to a much flatter structure which gives much poorer 'nests' for the cations to occupy. This partly explains the observed inability of such chains to complex except at higher ion concentrations.

The observed threefold screw symmetry of calcium mannuronate contrasts with the twofold guluronate, particularly since both are twofold in the acid form [146].

The chains of alternating mannuronate and guluronate show a few possible combinations with three or fourfold symmetry, but these are very heavily outnumbered by more open structures which cannot complex.

Consideration of the packing of such complexed chains suggests a structure resembling an egg-box (see fig. 40). Here each ion can complex with two chains, provided each is sufficiently bent, in accord with the results presented above. The structure will undoubtedly

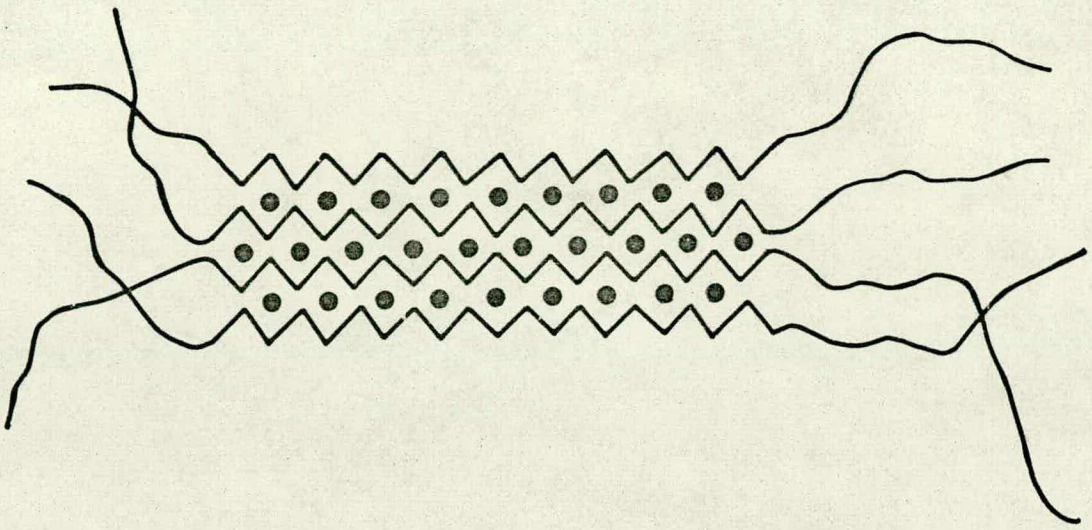


Fig. 40. The egg-box model.

reinforce in three dimensions, particularly for the cases of three-fold symmetry, where the simple picture must be elaborated.

In this model, the binding will be cooperative in the sense that the aggregates, once formed, will easily accept further strands, forming a compact three-dimensional structure.

5.3 CARRAGEENAN - THE COIL TO HELIX TRANSFORMATION

Carrageenans are gel-forming sulphated polysaccharides obtained from certain seaweeds. Their wide use industrially as food additives derives from their

texturing properties, the molecular and conformational basis of which is thus of considerable interest.

Rees and co-workers have shown that ι- and κ-carrageenan (VI) can exist in a double helical form in the solid state [30] and in solution [50]. The solution form undergoes a helix to random coil transition on being heated, which is reversed on cooling.

For any polymer, the existence of a regular helical structure is entropically unfavourable, since it implies that each linkage is restricted to the same conformation. For a helix to be stable, thus, there must be overriding enthalpy terms.

The most widely studied coil to helix transition is that of the polypeptide α-helix, which is a single helix. The enthalpic stability here was originally ascribed [19] to hydrogen bonding between the turns of the helix, but has subsequently been shown to derive from polar and van der Waals interactions as well. The hydrogen bond term in particular makes the transition cooperative, in the sense that the average number of stabilising hydrogen bonds per turn of helix increases with the number of turns. In this system there is a critical length of helix of n^* residues such that any larger one is stable with respect to the coil: i.e. if $\Delta G(n)$ is the free energy difference

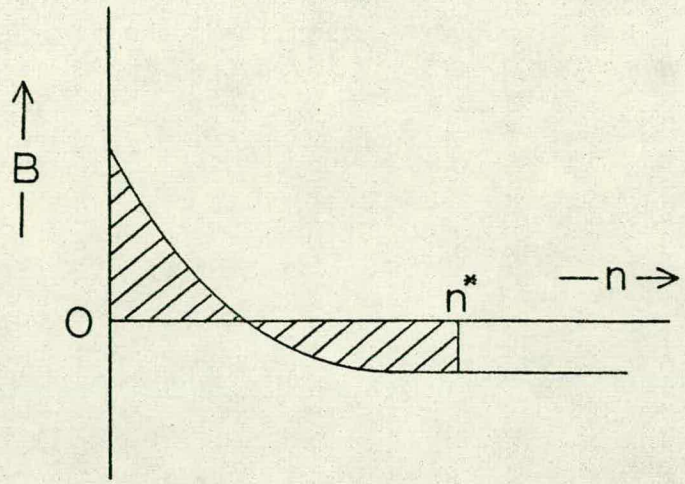


Fig. 41. The free energy (B) required to add the n th residue. The shaded areas are equal.

---

between n residues of coil and of helix, and

$B(n) = \Delta G(n) - \Delta G(n-1)$, the additional free energy for adding the nth helix residue, then B will vary as in fig. 41, and n^* will be such that

$$\Delta G\ (n^*) = \sum_{n=1}^{n^*} B\ (n) = 0 \qquad (1)$$

A treatment essentially similar to this has been applied to the α-helix by Brant [151].

Zimm and Bragg [152] have expressed this in more formal terms. They represent the chain as a sequence of Os and 1s, O representing a residue which is not H-bond

stabilised and 1, one that is. They assign a probability to any sequence as the product of the following terms:

i) unity for every 0;

ii) s for every 1 following a 1;

iii) σs for every 1 following μ or more 0s; and

iv) zero for every 1 following fewer than μ 0s.

Factor (iv) eliminates sequences of less than μ bonds: e.g. in the α-helix with 3 bonds per turn, a sequence of two missing bonds is deemed impossible.

The ratio 1:s then represents the relative probabilities of a residue existing as coil or helix, other than as the end residue, so that $s > 1$ represents a stable helix and vice versa. The quantity σ, which is often 10^{-2} or less, is the diminution of s for the residue at the 'growing' end of a helical region.

This treatment is clearly very simplified, since, for example, the transition from σs to s should obviously be gradual: nevertheless this model and developments of it have been of considerable use in helix-coil transition studies.

For carrageenan, where the transition is to a double helix, clearly the situation is rather different. For the first turn to form there is an additional, probably large, entropy term arising from the need to bring the two

chains together. In order, therefore, that the transition may occur at normal temperatures, there must exist a similarly large favourable enthalpy term for the helix. If ΔH and ΔS are the differences between coil and helix enthalpy and entropy then at the transition temperature $\Delta G = 0$, i.e. $\Delta H = T\Delta S$ and $T = \Delta H / \Delta S$. The sharpness of the transition is given by Schellmann's treatment [153]; thus if (c) and (h) are the concentrations of coil and helix,

$$K = (c)/(h) = \exp(-\Delta G/RT)$$
$$= \exp(-\Delta H/RT) \exp(\Delta S/R), \qquad (2)$$

whence $\partial K / \partial T = \Delta H / RT^2 . K$ (3)

The fraction existing as coil, f, is

$$f = (c) / (c) + (h) = K/1+K \qquad (4)$$

$$\therefore \quad \partial f / \partial T = 1/(1+K)^2 . \partial K / \partial T \qquad (5)$$

At the transition temperature, $K = 1$, thus

$$\partial f / \partial T = \Delta H / 4RT^2 \qquad (6)$$

so that the sharpness of the transition is enhanced if ΔH is large.

This treatment makes certain restricting assumptions, notably, that each molecule is either entirely coil or entirely helix, and does not exist with helical and coil regions along its length. It is also applicable only to monodisperse systems, since molecules of different lengths will have different values of ΔH. Such polydispersity, as occurs in most polymer samples, will manifest

itself as a broadening of the transition.

An attempt was made to calculate ΔH for carrageenan by the methods previously described. The molecule chosen for study was a completely unsulphated one. Although such a species has not yet been isolated, the series of ι-carrageenan (with 2 sulphate groups per disaccharide unit), κ-carrageenan (with 1) and furcellaran (with about 0.4) shows features suggesting that the stability of the helix increases with decreasing sulphation, notably in that the transition temperature increases and the calorimetric heat of the transition increases. In any case, because the sulphate groups occur on the outside of the helix, their absence is unlikely to affect greatly the conformation of the chain. Fig. 42 shows a photograph of a Beevers model of Rees' ι-carrageenan double helix, which has $n = 3$ (disaccharides) and $h = 8.67$ Å, showing 5 disaccharide residues on each strand. The strands are parallel but staggered.

The interaction energy was calculated for interaction of one residue on one strand with several on the other. It was found that (see fig. 42), taking residue 3 on strand 2, interactions with residues beyond 1 and 5 on strand 1 were negligible. Because the chains are staggered, the strand 2 residue was BA, where B represents the galactose and A the anhydrogalactose rings, whereas those on strand 1 were AB.

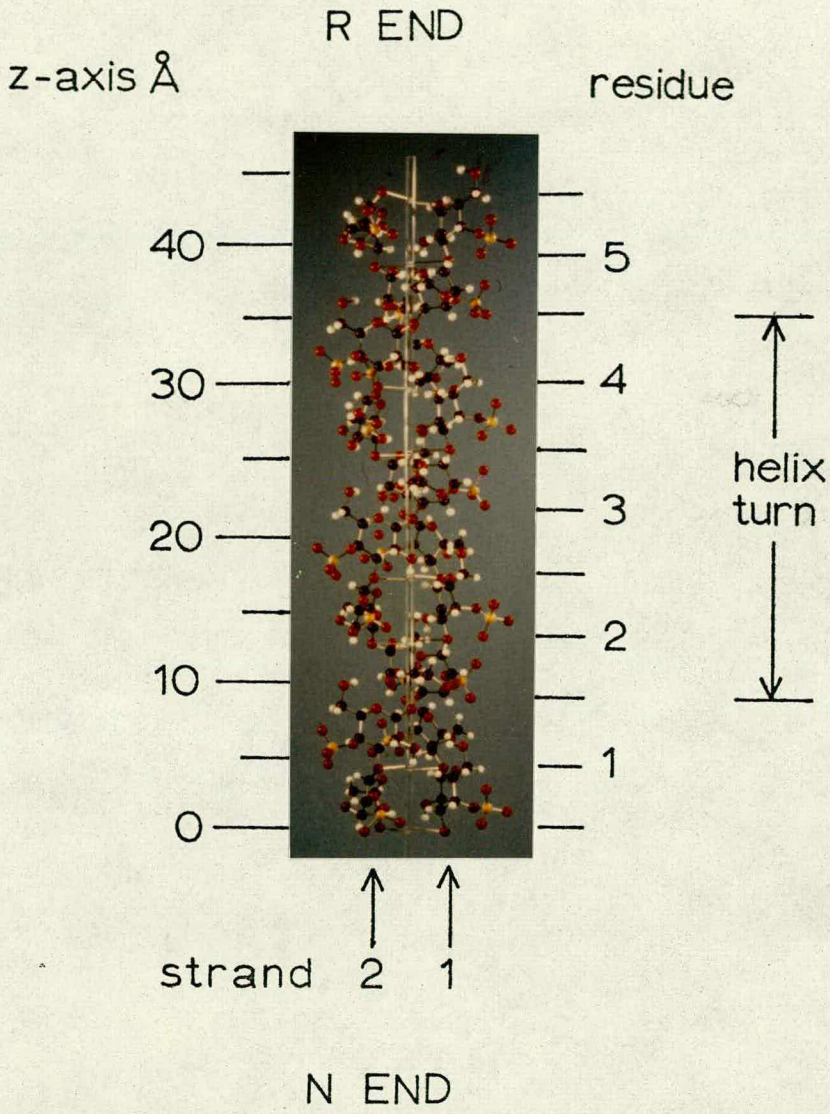


Fig. 42. Beevers model of the ι-carrageenan double helix, n = 3, h = 8.67 Å.

Coordinates for the disaccharide unit were taken from the x-ray work of Campbell [154]. Williamson [155] has noted from infra-red studies the existence of a hydrogen bond between H(O6) and O(2)' on the B residues, and to optimise this O(6) and H(O6) were fixed in the plane of C(5), C(6) and O(2)', thus fixing also H(6) and H(7), and H(O2)' was fixed in the plane of C(2)', O(2)' and O(6). The remaining unfixed hydrogen atoms, H(O4) on B and H(O2) on A were fixed in torsion minima 180° from C(5) and C(1) respectively, since their position would not seem to be of much significance. The energy calculation was then as before, using terms for van der Waals and polar interactions only, all torsional terms being constant.

The results are shown in table 19, from which it can be seen that the strands attract each other quite strongly. These energies will be increased by the hydrogen bond energy but much reduced by the fact that a helix strand will be less solvated than the coil because of the blocking effect of the other strand.

From table 19, the energy associated with a particular opposite pair of disaccharides may be calculated as the sum of their mutual interaction energy, plus half of their interaction energy with other residues; thus within the helix this is -13.4 kcal/mole of disaccharides and for

Table 19 CARRAGEENAN INTERSTRAND INTERACTION ENERGIES (kcal/mole of disaccharides)

Residue on strand 1	Interaction energy with residue on strand 2.						
	A			B			dis-dis
	polar	VdW	total	polar	VdW	total	total
1A.	0.00	0.00	0.00	0.01	0.00	0.01	0.00
B	0.00	0.00	0.00	0.02	-0.03	-0.01	
2A.	0.00	-0.02	-0.02	-0.39	-0.75	-1.14	-3.72
B	-0.04	-0.11	-0.15	-0.64	-1.77	-2.41	
3A.	0.16	-1.49	-1.33	-0.39	-2.87	-3.26	-10.26
B	-0.39	-2.87	-3.26	-0.64	-1.77	-2.41	
4A.	0.16	-1.49	-1.33	-0.04	-0.11	-0.15	-2.63
B	-0.39	-0.75	-0.14	0.02	0.03	0.01	
5A.	0.00	-0.02	-0.02	0.00	0.00	0.00	-0.02
B	0.01	0.00	0.01	-0.01	0.00	-0.01	

(Because of symmetry, certain of these terms are equal)

the last residues at the reducing and non-reducing ends respectively -12.1 and -11.6 kcal.

An additional factor which may favour either helix or coil arises from the freezing of the linkage conformation in the helix at a position which is not of average energy when adjacent residue interactions are considered as in chapter 4. If the helix linkage conformation is close to the minimum, this will favour the helix (since the coil mean energy will be somewhat higher than the minimum), otherwise it will favour the coil.

Disaccharide interactions were calculated for both linkages using the same coordinates as in the helix calculation and the usual potential function. In both cases, the helix linkage conformation is close to a local energy minimum but relatively remote from the global minimum which in each case is about 0.5 kcal/mole lower. Since the coil is likely to contain a substantial fraction of this global minimum conformation, this result is in accord with the observed considerable optical rotation change for the transition [50, 157]. For the 1,3 linkage, the double helix conformation is about 2.5 kcal/mole above the global minimum and for the 1,4 linkage, about 3.0 kcal/mole. It is difficult to estimate the mean energy from this source of the coil, but it is likely to be no more than 1 kcal/mole above the subsidiary minimum,

assuming as a worst case that the global minimum is largely precluded by longer range interactions.

Thus, by this effect the stability of the helix calculated above must be reduced by about 2 kcal/mole per linkage constrained to helix conformation, or 8 kcal/mole for a pair of disaccharides. The magnitude of this effect at the ends of the helical regions is not easily estimated, but may well be greater, in that linkages being changed to the double helix conformation may have to pass through higher energy conformations.

Since the available coordinates are from an unrefined crystal structure determination, the values of ϕ and ψ in the helix are somewhat uncertain. Arnott and Scott[158], however, have made a preliminary refinement of the structure and have estimated that neither value will change by more than 10°. In the extreme case, this would approximately half the effect noted here.

The conformational potential energy arising from interactions within and between the strands is thus estimated as about -5 kcal/mole of pairs of disaccharides within the helix or higher (less negative) at the ends.

The following factors must be taken into account in the consideration of the relative stabilities of the

forms and of the transition.

i) Against helix existence: solvation (a stabilising process) is greater in the coil.

ii) Against helix existence: the coil is entropically favoured.

iii) Against helix existence: the linkage conformation of the helix is not that giving lowest enthalpy for adjacent residue interactions.

iv) Against helix initiation: there is loss of entropy on bringing the strands together.

v) Against helix initiation: the first pair of residues have higher (i.e. smaller negative) enthalpy than subsequent ones (calculated above).

vi) For helix existence: van der Waals and(particularly) polar forces hold the strands together.

vii) For helix existence: hydrogen bonds hold the strands together (but note that unlike the α-helix case, this bond is perpendicular to the helix axis and thus does not contribute to cooperative helix formation).

It is difficult to do more than speculate on the relative importance of these factors.

Experimental work [50], which is complicated by the polydispersity of the samples used, shows typically a transition over about 30 to 50 K degrees, which displays hysteresis. Calorimetric studies [159], which give a direct

measurement of the net enthalpy change (including solvation) give for ι-carrageenan $\Delta H = 0.8$ kcal/mole, for κ-carrageenan, $\Delta H = 1.1$ kcal/mole and for agar, $\Delta H = 1.2$ kcal/mole. Agar is an unsulphated polysaccharide similar to carrageenans but with L anhydro residues.

In the terms of the Zimm and Bragg theory, σ is likely by virtue of term (iii) to be much less than unity, and s will vary markedly with temperature on account of the large entropic contributions to ΔG.

In conclusion, the thermodynamic balance is clearly complex and the field is open to the development of methods of measuring or calculating the factors involved. The present treatment makes no pretence to do more than suggest the sources of these factors.

5.4 THE XYLAN TRIPLE HELIX

The structure of β-1,3 xylan has been postulated by Atkins [160] to be triple helical, with $|n| = 6$ and $h = 3.06$ Å.

A similar calculation to that described in the preceding section was carried out for this polymer using Arnott and Scott coordinates. Because of the symmetrical nature of the system, interactions between only two strands need to be calculated.

The results are shown in table 20 and suggest that the right hand helix is slightly stabler. Considering each of each of the three strands, the energy of a monosaccharide residue remote from the ends of a strand is -6.0 kcal/mole of residues for the left-handed helix and -7.2 for the right-handed one. Each of these will be increased in magnitude by the energy of the hydrogen bonds formed between the O(2) atoms, the geometry and thus the energy of which is almost identical (but mirror-imaged) for the two helix forms.

On this basis, therefore, it would appear that neither structure is impossible, and that the right-handed helix is perhaps slightly favoured. Atkins comes to a similar conclusion on the basis of his x-ray work. Rao [161] has made a calculation using a Kitaigorodsky potential with no polar terms and concludes that the energy of each form is almost the same, but endorses Atkins' suggestion that the right-handed helix can form weak hydrogen bonds with water molecules which bridge the strands.

Until further evidence is available, therefore, it must be assumed that neither structure is ruled out.

5.5 B-AMYLOSE - SINGLE OR DOUBLE HELIX?

Reference has already been made to the controversy regarding the conformation of B-amylose.

TABLE 20

Xylan triple helix interstrand interaction energies.

Interaction (kcal/mole) of one residue with the residue on another strand displaced m residues along towards the reducing end of the chain.

m	Left hand helix	Right hand helix
-5	-0.01	-0.03
-4	-0.02	-0.07
-3	-0.06	-0.26
-2	-0.15	-2.12
-1	-0.33	-3.10
0	-1.23	-0.70
1	-2.45	-0.44
2	-1.49	-0.15
3	-0.15	-0.13
4	-0.05	-0.10
5	-0.02	-0.05

In section 4.3 the results of energy calculations on the α-1,4 linkage are presented, and the proposed conformations, both single and double helix, are shown on the energy maps.

The double helical structure, with n = -6 and h = 3.46 Å is, as previously stated, geometrically impossible without some distortion of the A & S geometry. The necessary distortion was calculated to be either a reduction of the glycosidic angle by 4° to 113°, giving the required n and h at $\phi = 162^{\circ}$ and $\psi = 143^{\circ}$, or an increase in the 'virtual bond angles' (see section 4.2; μ_2 and μ_3 in fig. 25) by 2° each, yielding $\phi = 164^{\circ}$, $\psi = 144^{\circ}$. This last distortion, if distributed among the contributory bond lengths and angles, is likely to be well within the precision limits of the A & S coordinate set. These points, and the nearest approximation without distortion are very close to the calculated minimum energy linkage conformation, and for this reason, this structure was further investigated.

Rees [162] has described a procedure ('Modelbuild II') for examining the feasibility of double helical structures, using hard-sphere criteria. This consists of fixing one helical strand and moving another residue from a position initially coincident with a residue on this strand, parallel to the helix axis (the d direction) and rotationally about the helix axis (the θ direction). A hard-sphere map in

d and θ is then produced, representing all possible double helical structures with parallel strands, and another similarly yielding antiparallel structures.

The position of highest h with n = -6 was accurately located (see section 4.2) as $\phi = 156.92^{\circ}$, $\psi = 163.49^{\circ}$, h = 2.999 Å and using Rees' procedure it was found that a double helix is possible only for parallel chains where d = 9 Å and $\theta = 0^{\circ}$, i.e. with the two strands symmetrically placed and corresponding atoms in the same plane perpendicular to the helix axis (see fig. 43).

This structure can be seen from models to have a 'hole' along the helix axis, so that stretching to the observed h = 3.46 Å will ease the interstrand interaction without introducing any major clashes across the helix axis.

From this work no clear preference can be established between single and double helical structures. However, if the strain required to extend the projected residue height by about 0.5 Å is not too great, the double helical structure would appear to be the one with linkage conformation closest to the calculated minimum of enthalpy.

Several other helical amylose structures have been observed, notably the triacetate [65] and the complex with potassium bromide[163]. Values of n, h, ϕ and ψ for these,

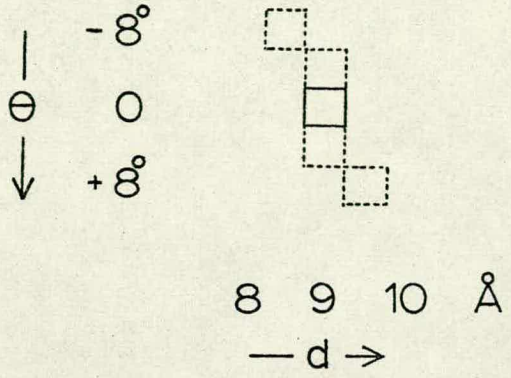


Fig. 43. Hard sphere map of amylose double helix, showing fully (solid line) and partially (broken line) allowed regions in d, displacement parallel to the helix axis, and θ, rotation about the helix axis, defining the second strand relative to the first.

assuming A & S geometry, are given in table 21. It is of note that this geometry, unlike that of Blackwell et al[25] permits all these forms to exist, as well as the six- and seven-fold complexed structures. Clearly, however, the helix parameters in this system are very sensitive to small changes in the ring geometry.

5.6 CONCLUSION

In 1969, when the work described in this thesis was started, almost all published calculations on carbohydrate conformation were on the basis of the hard sphere approach. Since then, several attempts, including this one, have been made to calculate energies, and already these are beginning to provide solutions to structural and conformational problems.

The value of these calculation methods lies in their ability to analyse the contributory factors determining the stability of systems, and thereby to give insights at a molecular level which may be used to direct further experimental work. The scope for further work is almost unlimited, both in the application of the present methods and in the development of refinements and extensions to cover a wider range of molecules and systems.

TABLE 21

Calculated ϕ and ψ values for helical amylose structures.

Structure	Reported		Calculated	
	n	h(Å)	ϕ	ψ
Triacetate	-14/3	3.75	134°	151°
KBr complex	-4	4.03	122°	139°

APPENDICES

A 1.1 APPENDIX 1 - FUNCTION OPTIMISATION

The problem frequently arises in conformational work of finding the optimum or minimum value of a function of many variables, $F(\underline{x})$, $\underline{x} = x_1, x_2, \ldots$ Many algorithms exist [118] for doing this, most of which require a starting point $\underline{x}_o$ from which they iteratively descend to a minimum of F, which may or may not be the global minimum if F has any subsidiary local minima. Some require a procedure for calculation the derivate of F with respect to each x at any point. These have the advantage in general of requiring fewer function evaluations to achieve a given accuracy, but the disadvantage that the derivatives may not be easily calculable.

A full review of available methods is outwith the scope of this thesis and is available elsewhere [118], but a brief mention follows of several methods found useful in this work.

A very powerful method requiring derivatives has been described by Davidon [164], and in a modified form by Fletcher and Powell [165]. Powell [166] subsequently adapted this procedure to operate without derivatives in a method which is reliable and rapid. More recently, Stewart [167] has published an even better version, an algorithm for which has been given by Lill [168]. This was coded in Fortran

and found to give excellent convergence, but to be troublesome for certain combinations of function and gradient values.

A less efficient but very robust method has been proposed by Nelder and Mead [169]. This 'simplex' method has been extensively used and is discussed briefly in the next section.

The problem of finding a global minimum amongst several local ones has been little considered. Crippen and Scheraga [170] have however devised a sophisticated 'spotting' algorithm which does this effectively and have used it to find the minimum energy conformation of tetraglycine.

For general use, when the number of function evaluations is not unduly critical, thus, the methods of Powell [165] or Nelder and Mead [169] can be recommended. For work where the function takes a long time to evaluate it may be worth trying Stewart's algorithm [167].

A 1.2 SIMPLEX MINIMISATION

Nelder and Mead's [169] simplex method has been used frequently in this work. It has the advantage of being simple to use and understand. A brief description follows.

A simplex is an array of n+1 points in n-dimensional space, for example a (not necessarily regular) tetrahedron in 3-dimensional space. The method takes an initial simplex in the space of the variables of the function under consideration and modifies it iteratively until it contracts on the function minimum. Three modifications are used, reflection, contraction and expansion.

Reflection consists of reflecting the simplex vertex of highest function value in the centroid of all the other vertices (fig. 44). Contraction follows reflection in certain circumstances, and is a moving of the same point half way back towards the centroid of the others (fig. 45). Expansion also follows reflection and consist of moving the same point twice as far from the centroid of the others (fig. 46). Each of these operations is considered as a success if the moved point has its function value lower than all the others, or to be a failure if not. Fig. 47 shows the logic of the process. The salvage operation required for a failed contraction occurs only rarely in practice and comprises moving every other point halfway towards the best point.

The criterion for termination of the iterations used is twofold: the procedure terminates with the centroid of the simplex when two successive iterations (or groups

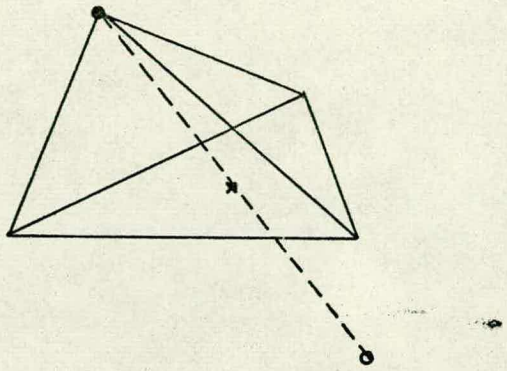


Fig 44 Reflection

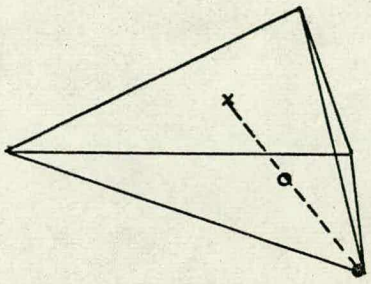


Fig 45 Contraction

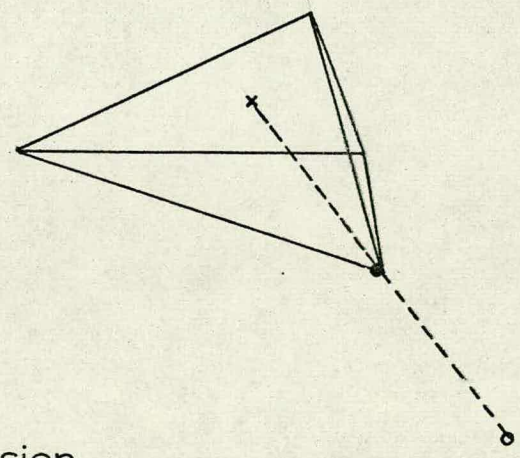


Fig 46 Expansion

Simplex operations in 3-space.
● is the point being moved, which is initially (fig. 44) that of highest function value.
X is the centroid of all the other points.
O is the new position of the moved point.

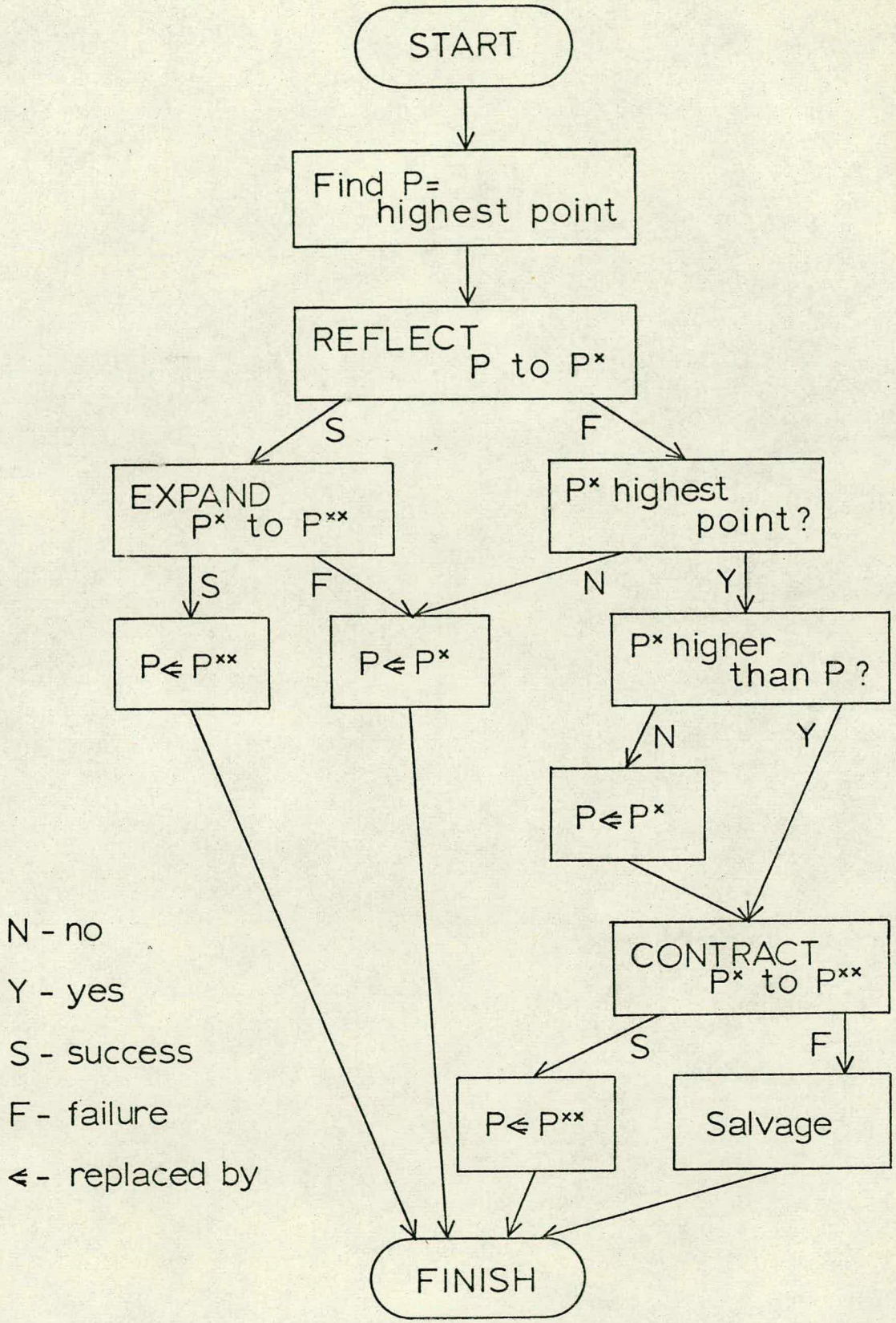


Fig 47 Simplex iteration

of iterations) show 'convergence' to points 'sufficiently close' together or a specified number of iterations has been completed. 'Convergence' is defined as when the root mean square distance from each point to the centroid is less than a preset value, and points are 'sufficiently close' if within that distance.

A 2. APPENDIX 2 - RANDOM NUMBERS

The random numbers used in the Monte-Carlo methods described earlier were in fact pseudorandom numbers generated by the congruence method [171].

If x_j is one of a pseudorandom sequence then x_{j+1} is given by

$$x_{j+1} = (ax_j) \text{ modulo } n$$

where a and n are chosen according to established principles. For this work the values used [172] were

$a = 65\ 539, \quad n = 2\ 147\ 483\ 648\ (= 2^{31})$.

A 3. APPENDIX 3 - GEOMETRICAL MANIPULATIONS

Several essentially geometrical operations occur frequently in this thesis: they are gathered together here. In this appendix P_1, P_2 etc. are points with coordinates (x_1, y_1, z_1), (x_2, y_2, z_2) etc; $\underline{r}_{ij}$ is the vector $\overrightarrow{P_iP_j}$, magnitude r_{ij}, components x_{ij}, y_{ij} and z_{ij}.

(i) Calculation of angles

The angle $P_iP_jP_k$, denoted θ_{ijk}, is most easily derived from the identity $\underline{r}_{ji}.\underline{r}_{jk} = r_{ji}r_{jk}\cos\theta_{ijk}$, whence

$$\theta_{ijk} = \arccos\left((x_{ji}x_{jk}+y_{ji}y_{jk}+z_{ji}z_{jk})/r_{ji}r_{jk}\right) \qquad (1)$$

(ii) Rotation about a coordinate axis.

If P_i is rotated by ϕ clockwise, looking from $+\infty$, about the x axis to become P_i', then

$$\begin{bmatrix} x_i' \\ y_i' \\ z_i' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \qquad (2)$$

or $\underline{r}_{Oi}' = \underline{\underline{T}}_x(\phi) \cdot \underline{r}_{Oi}$

Similarly,

$$\underline{\underline{T}}_y(\phi) = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 1 & 0 & 1 \\ \sin\phi & 0 & \cos\phi \end{bmatrix}, \quad \underline{\underline{T}}_z(\phi) = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(iii) Alignment upon axes

It frequently is required that a set of points be rotated about the origin such that a specified one lies on a given semi-axis. For example (see fig. 48) if P_i is required to lie on the positive x-axis, two rotations of the system are required, by η about the y-axis followed by ζ about the z-axis.

Define $A(y,x) \equiv \lceil \arctan(y/x), \quad x > 0;$ $\arctan(y/x) + \text{sign}(y).\pi, \quad x < 0;$ $\text{sign}(y).\pi/2,$ $x = 0$ and $y \neq 0$, undefined if $x = y = 0 \rceil$, arc tan being taken in the range $-\pi < \arctan(p) \leqslant \pi$. This function is supplied by many computer systems (two argument arctangent).

Then $\eta = A(-z_i, x_i)$ and $\zeta = A(y_i, \sqrt{x_i^2 + z_i^2})$ and the transformation matrix required is thus $\underline{\underline{R}}$, where

$$\underline{\underline{R}} \equiv \begin{bmatrix} \cos\zeta\cos\eta & \sin\zeta & -\cos\zeta\sin\eta \\ -\sin\zeta\cos\eta & \cos\zeta & \sin\zeta\sin\eta \\ \sin\eta & 0 & \cos\eta \end{bmatrix} \qquad (3)$$

(iv) Calculation of dihedral angles

The following procedure determines the dihedral angle between the planes defined by $P_iP_jP_k$ and $P_jP_kP_l$.

Let it be β_{ijkl}.

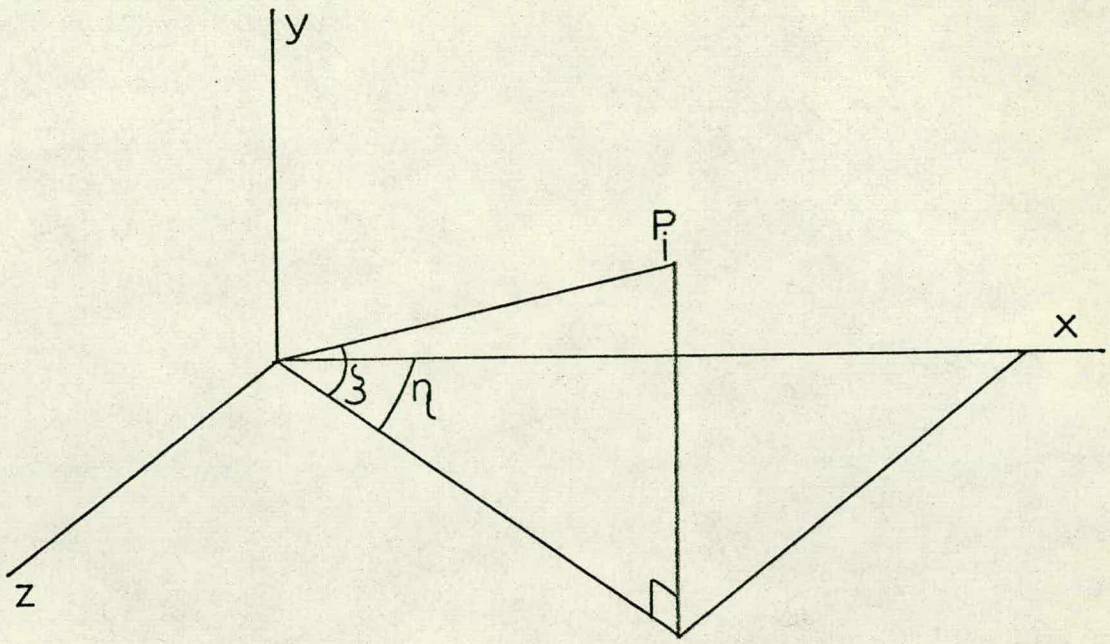


Fig. 48. Rotation angles required to align OP_1 with the x-axis.

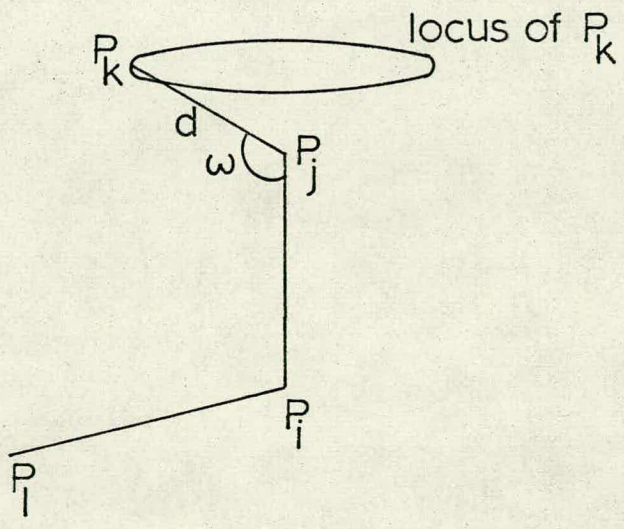


Fig. 49. The floating atom problem.

Let $\eta = A\,(-z_{jk},\ x_{jk})$, A as above.

$$\zeta = A\,(y_{jk},\ \sqrt{x_{jk}^2 + z_{jk}^2})$$

whence matrix $\underline{\underline{R}}$ (equation 3).

Let $\underline{r}_{ji}' = \underline{\underline{R}} \cdot \underline{r}_{ji}$ and $\underline{r}_{kl}' = \underline{\underline{R}} \cdot \underline{r}_{kl}$.

Then $\beta_{ijkl} = A\,(z_{kl}',\ y_{kl}') - A\,(z_{ji}',\ y_{ji}')$.

Angle β will be positive if, when looking from P_j to P_k, P_l is $|\beta|$ clockwise from P_i.

(v) Rotation of a point about an arbitrary axis.

This 'floating atom' problem is that of placing (see fig. 49) a point P_k at a distance d from P_j such that angle θ_{ijk} is w and dihedral angle β_{lijk} is v. The problem usually relates to constant d and w and varying v, accordingly, the following procedure

a) sets up an array of 15 quantities X_i, i = 1 to 15 from the positions of P_i, P_j and P_l, and d and w, and

b) uses it, repetitively if required, to generate P_k from X and v.

Let X_{11}, X_{12} and X_{13} be x_j, y_j, z_j respectively.

If $x_{ij} = z_{ij} = 0$, let $\eta = 0$; otherwise let $\eta = A\,(z_{ij},\ x_{ij})$ (see above for A).

Let $\zeta = A(y_{ij},\ \sqrt{x_{ij}^2 + z_{ij}^2})$

Let $X_{15} = d \sin (\pi - w)$.

and $X_{14} = d\, r_{ij} \cos (\pi - w)$.

Let $\begin{bmatrix} X_1 & X_4 - X_7 \\ X_2 & X_5 - X_8 \\ -X_3 & X_6 & X_9 \end{bmatrix} = \underline{\underline{R}}$ from η and ζ (eqn. 3)

Let $a = -\cos\eta \sin\zeta\, x_{il} + \cos\zeta\, y_{il} - \sin\eta \sin\zeta\, z_{il}$

and $b = -\sin\eta\, x_{il} + \cos\eta\, z_{il}$

and finally $X_{10} = A\,(b,a)$.

We now have all fifteen elements of X.

For any v, then, let $u = v + X_{10}$,

let $p_1 = X_{14}$, $p_2 = X_{15} \cos u$, $p_3 = X_{15} \sin u$,

and then

$$\begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & X_3 \\ X_4 & X_5 & X_6 \\ X_7 & X_8 & X_9 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} + \begin{bmatrix} X_{11} \\ X_{12} \\ X_{13} \end{bmatrix}$$

A 4. APPENDIX 4 - COMPUTING DETAILS

The programs used in this work were mostly written in Fortran IV. They were run variously on the IBM 360/50 and 370/155 at the Edinburgh Regional Computing Centre and the 360/50 and 360/65 at Unilever Computing Services Watford Data Centre under the G and H compilers, and on the ERCC ICL 4/75 under the ERCOS Edinburgh Fortran compiler. The Honeywell timesharing terminal service on a GE 265 computer using enhanced Fortran II was also found invaluable for preliminary programming and for geometrical work.

Running times for the two long programs detailed in the preceding chapters on the 360/50 under H compilation were, for the monosaccharide program (chapter 3) about 9 mins per conformation or about $2\frac{1}{2}$ hours for all 16 and for the disaccharide program (chapter 4) about 5 mins per ϕ/ψ combination for which the averaging phase was entered, giving a total time of about 12 hours for maltose.

REFERENCES

REFERENCES

1. Reeves, R.E., J.Amer.Chem.Soc., 71, 215, (1949).
2. Haworth, W.N., 'The Constitution of the Sugars', Arnold, London, (1929).
3. Barton, D.H.R., Experientia, 6, 316, (1950).
4. for example, Eliel, E.L., Allinger, N.L., Angyal, S.J., Morrison, G.A., 'Conformational Analysis', Wiley, NY, (1965), page 4.
5. Barton, D.H.R., Cookson, R.C., Quart. Rev., 10, 44,(1956).
6. commencing with Eliel, E.L., Pillar, C., J.Amer.Chem.Soc. 77, 3600, (1955).
7. Eliel, E.L., Gilbert, E.C., J.Amer.Chem.Soc., 91, 5487, (1969).
8. Reeves, R.E., Jung, J.R. jr., J.Amer.Chem.Soc., 71, 209, (1949).
9. Reeves, R.E., J.Amer.Chem.Soc., 71, 212, (1949).
10. ibid., 71, 215, (1949).
11. ibid., 72, 1499, (1950).
12. Angyal, S.J., Austral.J.Chem., 21, 2737, (1968).
13. Angyal, S.J., Dawes, K., Austral.J.Chem., 21, 2747,(1968).
14. Durette, P.L., Horton, D., Adv.Carbohydrate Chem. Biochem., 26, 49, (1971).
15. Arnott, S., Scott, W.E., J.Chem.Soc. Perkin II, 324, (1972).
16. Marchessault, R.H., Sarko, A., Adv.Carbohydrate Chem., 22, 421, (1967).

17. Watson, J.D., Crick, F.H.C., Nature, 171, 737, (1953).

18. ibid., 171, 964, (1953).

19. Pauling L., Corey, R., Branson, H., Proc. Nat. Acad.Sci., U.S.A., 37, 205, (1951).

20. Rees, D.A., Scott, W.E., J.Chem.Soc.(B), 469, (1971).

21. Rees, D.A., in MTP International Review of Science: Organic Chemistry series 1, 7, vol. ed. G.O. Aspinall, Butterworth's, London, (1972).

22. Meyer, K.H., 'Natural and Synthetic High Polymers', Interscience, N.Y., (1950).

23. Settineri, W.J., Marchessault, R.H., J.Polymer Sci., C11, 235, (1965).

24. French, A.D., Zashow, B., Chem.Comm. 41, (1972).

25. Blackwell, J., Sarko, A., Marchessault, R.H., J.Mol.Biol., 42, 379, (1969).

26. French, D., in 'Symposium on Foods: Carbohydrates and their Roles', ed. H.W. Schulz, A.V.I., Westport, Connecticut, (1969), page 26.

27. Atkins, E.D.T., Parker, K.D., J.Polymer Sci., C28, 69, (1969).

28. Atkins, E.D.T., Sheehan, J.K., Nature New Biol., 235, 253, (1972).

29. Dea, I.C.M., Moorhouse, R., Rees, D.A., Arnott, S., Guss, J.M., Balazs, E.A., in preparation.

30. Anderson, N.S., Campbell, J.W., Harding, M.M., Rees,D.A., Samuel, J.W.B., J.Mol.Biol., 45, 85, (1969).

31. Lemieux, R.U., Kellnig, R.K., Bernstein, H.J., Schneider, W.G., J.Amer.Chem.Soc., 80, 6098, (1958).

32. Lemieux, R.U., Stevens, J.D., Canad.J.Chem., 43, 2059, (1965).

33. ibid. 44, 249, (1966).

34. Lemieux, R.U., Pure Appl.Chem., 27, 527, (1971).

35. Rudrum, M., Shaw, D.F., J.Chem.Soc., 52, (1965).

36. Angyal, S.J., Angew.Chem.Internat.Edn., 8, 157, (1969).

37. Perlin, A.S., Casu,B., Koch, H.J., Canad.J.Chem., 48, 2596, (1970).

38. Stoddart, J.F., 'Stereochemistry of Carbohydrates', Wiley, N.Y., (1971), page 137.

39. Karplus, M.,J.Chem.Phys., 30, 11, (1959).

40. Casu, B., Reggiani, M., Gallo, G.G., Vigevani, A., Tetrahedron, 22, 3061, (1966).

41. ibid., 24, 803, (1968).

42. ibid. Carbohydrate Res., 12, 157, (1970).

43. Dorman, D.E., Roberts, J.D., J.Amer.Chem.Soc., 93, 4463, (1971).

44. Hudson, C.S., J.Amer.Chem.Soc., 31, 66, (1909).

45. Kauzmann, W., Eyring, H., J.Chem.Phys., 9, 41, (1941).

46. Whiffen, D.H., Chem. and Ind., 954, (1956).

47. Brewster, J.H., J.Amer.Chem.Soc., 81, 5475, (1959), and following papers.

48. Rees, D.A., J.Chem.Soc.(B), 877, (1970).

49. Rees, D.A., Scott, W.E., Williamson, F.B., Nature, 227, 390, (1970).

50 McKinnon, A.A., Rees, D.A. Williamson, F.B., Chem.Comm., 701, (1969).

51. Dea, I.C.M., McKinnon, A.A., Rees D.A., J.Mol.Biol., 68, 153, (1972).

52. Listowsky, I., Avigad, G., Englard, S., Carbohydrate Res., 8, 205, (1968).

53. Morris, E.R., Rees, D.A., Sanderson, G.R., in preparation.

54. Thom, D., Ph.D. Thesis, Univ. of Edinburgh, (1972).

55. Grant, G.T., Ph.D. Thesis, Univ. of Edinburgh, (1972).

56. Stone, A.L., in 'Structure and Stability of Biological Macromolecules', eds. G. Fasman, S. Timasheff, Dekker, N.Y., (1969).

57. Morris, E.R., Sanderson, G.R., in 'New Techniques in Biophysics and Cell Biology', eds. R. Pain, B. Smith, Wiley, London, (1972).

58. Spedding, H., Adv. Carbohydrate Chem., 19, 23 (1964).

59. Tipson, R.S., 'Infrared Spectroscopy of Carbohydrates', N.B.S. monograph 110, Washington, (1968).

60. Kuhn, L.P., J.Amer.Chem.Soc., 74, 2492, (1952).

61. ibid., 76, 4323, (1954).

62. ref. 38, chap.4.

63. Williams, J.E., Stang, P.J., Schleyer, P.von R., Ann. Rev. Phys. Chem., 19, 531, (1968).

64. Pauling, L., 'The Nature of the Chemical Bond', Cornell, N.Y., 3rd ed. (1960).

65. Sarko, A., Marchessault, R.H., J.Amer.Chem.Soc., 89, 6454, (1967).

66. Rees, D.A., Skerrett, R.J., J.Chem.Soc.(B), 189, (1969).

67. Rees, D.A., Scott, W.E., Chem.Comm., 1037, (1969).

68. Pitzer, K.S., Catalano, E., J.Amer.Chem.Soc., 78, 4844, (1956).

69. Brant, D.A., Flory, P.J., J.Amer.Chem.Soc., 87, 663, (1965).

70. Scott, R.A., Scheraga, H.A., J.Chem.Phys., 42, 2209, (1965).

71. Goebel, C.V., Dimpfl, W.L., Brant, D.A., Macromolecules, 3, 644, (1970).

72. Kitaigorodsky, A.I., Tetrahedron, 14, 230, (1961).

73. Rao, V.S.R., Yathindra, N., Sundararajan, P.R., Biopolymers, 8, 325, (1969).

74. Whittington, S.G., Biopolymers, 10, 1481, (1971).

75. Whittington, S.G., Glover, R.M., Macromolecules, 5, 55, (1972).

76. Allinger, N.L., Miller, M.A., Vancantledge, F., Hirsch, J.A., J.Amer.Chem.Soc., 89, 4345, (1967).

77. De Santis, P., Giglio, E., Liquori, A.M., Ripamonti, A., Nature, 206, 456, (1965).

78. Rao, V.S.R., Sundararajan, P.R., Ramachandran, G.N., Ramakrishnan, C., in 'Conformation of Biopolymers', ed. G.N. Ramachandran, Academic, London,(1967), 2, 721.

79. Venkatchalam, C.N., Ramachandran, G.N., in ref. 78, 1, 83.

80. Skerrett, R.J., Ph.D. Thesis, Univ. of Edinburgh, (1968).

81. Leach, S.J., Nemethy, G., Scheraga, H.A., Biopolymers, 4, 369, (1966).

82. Scott, R.A., Scheraga, H.A., J.Chem.Phys. 45, 2091, (1966).

83. Ooi, T., Scott, R.A., Vanderkooi, G., Scheraga, H.A., J.Chem.Phys., 46, 4410, (1967).

84. Del Re, G., Theor.Chim.Acta., 1, 188, (1963).

85. French, F.A., Rasmussen, R.S., J.Chem.Phys., 14, 389, (1946).

86. Scheraga, H.A., Adv.Phys.Org.Chem., 6, 103, (1968).

87. Allen, L.C., Chem.Phys.Letters, 2, 597, (1968).

88. Magnasco, V., Musso, G.F., Chem.Phys.Letters, 9, 433, (1971).

89. Lippincott, E.R., Schroeder, R., J.Chem.Phys. 23, 1099, (1955).

90. Schroeder, R., Lippincott, E.R., J.Phys.Chem., 61, 921, (1957).

91. Poland, D., Scheraga, H.A., Biochemistry, 6, 3791, (1967).

92. Gibson, K.D., Scheraga, H.A., Biopolymers, 4, 709,(1966).

93. Rao, V.S.R., Vijayalakshmi, K.S., Sundararajan, P.R., Carb. Res., 17, 341, (1971).

94. Pullman, B. Internat.J.Quantum Chem., 319, (1971).

95. Giacomini, M., Pullman, B., Maigret, B., Theor.Chim.Acta., 19, 347, (1970-.

96. Saran, A., Govil, G., Indian J.Chem., 9, 1095, (1971).

97. Blann, W.G., unpublished results.

98. Brant, D.A., Miller, W.G., Flory, P.J., J.Mol.Biol. 23, 47, (1967).

99. Del Re, G., J.Chem.Soc., 4031, (1958).

100. ibid., in 'Electronic Aspects of Biochemistry', ed. B. Pullman, Academic, NY, (1964), page 221.

101. see for example Fox, L., Mayers, D.F., 'Computing Methods for Scientists and Engineers', Clarendon, Oxford, (1968), chapter 5.

102. Zhdanov, Yu.A., Minkin, V.I., Dorofeenko, G.N., Ostroumov, Yu.A., Malisheva, E.N., Proc. 4th All-Union Conference on the Chemistry and Biochemistry of Carbohydrates, Moscow, (1969).

103. Vijayalakshmi, K.S., Rao, V.S.R., Carbohydrate Res., 22, 413, (1972).

104. Brant, D.A., Flory, P.J., J.Amer.Chem.Soc., 87, 2791, (1965).

105. Gibson, K.D., Scheraga, H.A., Proc.Nat.Acad.Sci. U.S.A., 58, 420 and 1317, (1967).

106. Sundararajan, P.R., Rao, V.S.R., Tetrahedron,24, 289,(1968).

107. Sundararajan, P.R., Ph.D. Thesis, Univ.of Madras (1969).

108. Ramachandran, G.N., Srinivasan, R., Indian J.Biochem., 7, 95, (1970).

109. Kemp, J.D., Pitzer, K.S., J.Amer.Chem.Soc., 59, 276, (1937).

110. Wilson, E.B., Proc.Nat.Acad.Sci. U.S.A., 43, 816, (1957).

111. Pauling, L., Proc.Nat.Acad.Sci., U.S.A., 44, 211, (1958).

112. ref. 64, page 131.

113. Oosterhoff, L.J., Pure Appl.Chem., 25, 563, (1971).

114. Pitzer, K.S., Gwinn, W.D., J.Chem.Phys., 10, 428,(1942), and references therein.

115. Kistiakowsky, G.B., Lacher, J.R., Stitt, F., J.Chem.Phys., 7, 289, (1939).

116. Miller, D.J., Progr.Stereochem., 3, 138, (1962).

117. see for example Hill, T.L., 'Introduction to Statistical Mechanics', Addison Wesley, Reading, Mass., (1960), chapter 1.

118. Kowalik, J., Osborne, M.R., 'Methods for Unconstrained Optimisation Problems', Elsevier, N.Y. (1968).

119. Fluendy, M.A.D., Smith, E.B., Quart.Rev. 16, 241, (1962).

120. Lowry, G.G., (ed.), 'Markov Chains and Monte-Carlo Calculations in Polymer Science', Dekker, N.Y. (1970).

121. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., J.Chem.Phys., 21, 1087, (1953).

122. Sundararajan, P.R., Marchessault, R.H., Biopolymers, 11, 829, (1972).

123. ref. 38, page 90.

124. Hutchins, R.O., Kopp, L.D., Eliel, E.L., J.Amer.Chem.Soc., 90, 7174, (1968).

125. Wolfe, S., Rauk, A., Tel, L.M., Csizmadia, I.G., J.Chem.Soc. B., 136, (1971).

126. Miyazawa, T., J.Polymer Sci., 55, 215, (1961).

127. Rees, D.A., Skerrett, R.J., Carbohydrate Res., 7, 334, (1968).

128. Manor, P.C., Saenger, W., Nature, 237, 392, (1972).

129. Saenger, W., private communication to Prof. D.A. Rees.

130. Brant, D.A., Dimpfl, W.L., Macromolecules, 3, 655, (1970).

131. Quigley, G.J., Sarko, A., Marchessault, R.H., J.Amer.Chem.Soc., 92, 5834, (1970).

132. Chu, S.S.C., Jeffrey, G.A. Acta Cryst. 23, 1038, (1967).

133. Hybl, A., Rundle, R.E., Williams, D.G., J.Amer.Chem.Soc. 87, 2779, (1965).

134. Yathindra, N., Rao, V.S.R., Biopolymers, 9, 783, (1970).

135. Sathyanarayana, B.K., Rao, V.S.R., Biopolymers, 10, 1605, (1971).

136. Chu, S.S.C., Jefirey, G.A., Acta Cryst. B., 24, 830, (1968).

137. Fries, D.C., Rao, S.T., Sundaralingam, M., Acta Cryst.B., 27, 994, (1971).

138. Ham, J.T., Williams, D.G., Acta Cryst. B., 26, 1373, (1970).

139. Morris, E.R., Rees, D.A., Thom. D., in preparation.

140. Grant, G.T., Morris, E.R., Rees, D.A., Smith, P.J.C., Thom, D., in preparation.

141. Gould, R.O., Gould, S.E.B., Rees, D.A., Scott W.E., in preparation.

142. Williams, R.J.P., private communication to Prof. D.A. Rees.

143. Bugg, C.E., Cook, W.J., J.Chem.Soc.Chem. Comm., 727, (1972).

144. Wight, A.W., Ph.D. Thesis, Univ. of Edinburgh, (1972).

145. Furberg, S., Jammer, H., Mostad, A., Acta Chem.Scand., 17, 2444, (1963).

146. Mackie, W., Biochem. J., 125 89P, (1971).

147. Rees, D.A., Wight, A.W., J.Chem.Soc.B., 1366, (1971).

148. Palmer, K.J., Hartzog, M.B., J.Amer.Chem.Soc., 67, 2122, (1945).

149. Kohn, R., Furda, J., Haug, A., Smidsrød, O., Acta Chem.Scand., 22, 3098, (1968).

150. Haug, A., Smidsrød, O., Acta Chem.Scand., 24, 267,(1970).

151. Brant, D.A., Macromolecules, 1, 291, (1968).

152. Zimm, B.H., Bragg, J.K., J.Chem.Phys., 31, 526, (1959).

153. Schellmann, J.A., Compt.Rend.Trav.Lab.Carlsberg, Ser.Chim., 29, 230, (1955).

154. Campbell, J.W., Ph.D. Thesis, Univ. of Edinburgh,(1969).

155. Williamson, F.B., Ph.D. Thesis, Univ. of Edinburgh, (1970).

156. Rees, D.A., Adv.Carbohydrate Chem.Biochem., 24, 267, (1969).

157. McKinnon, A.A., Ph.D. Thesis, Univ. of Edinburgh, (1972).

158. Arnott, S., Scott, W.E., private communication to Prof. D.A. Rees.

159. Reid, D.S., unpublished work.

160. Atkins, E.D.T., Parker, K.D., Preston, R.D., Proc.Roy.Soc. B., 173, 209, (1969).

161. Sathyanarayana, B.K., Rao, V.S.R., Carbohydrate Res., 15, 137, (1970).

162. Rees, D.A., appendix to ref. 30.

163. Jakobs, J.J., Bumb, R.R., Zaslow, B., Biopolymers, 6, 1659, (1968).

164. Davidon, W. C., USAEC Res.& Dev. Rpt. ANL5990 (rev.), (1959).

165. Fletcher, R., Powell, M.J.D., Comp.J., 6, 163, (1963).

166. Powell, M.J.D., Comp.J., 7, 303, (1965).

167. Stewart, G.W., J.Ass.Comput.Mach., 14, 72, (1967).

168. Lill, S.M., Comp.J., 13, 111, (1970); important corrections, ibid., 14, 106 and 214, (1971).

169. Nelder, J.A., Mead, R., Comp.J. 7, 308, (1965).

170. Crippen, G.M., Scheraga, H.A., Arch.Biochim.Biophys., 144, 453, (1971).

171. Abramowitz, M., Stegun, I.A., 'Handbook of Mathematical Functions', Dover, N.Y., (1964), page 949.

172. International Business Machines Corporation, 'Random Number Generating and Testing', manual C20-8011, I.B.M., White Plains, N.Y., (undated).